Université Paris I Panthéon-Sorbonne
U.F.R. de Philosophie

THÈSE
présentée publiquement le 15 décembre 2007
par

Isabelle Drouet

en vue de l'obtention du
Doctorat de l'Université Paris I
en
Philosophie

Causalité et probabilités : réseaux bayésiens, propensionnisme

---

Directeur de thèse :

M. Jacques Dubucs	Directeur de recherche au CNRS, Directeur de l'IHPST

Jury :

M. Donald Gillies	Professeur à University College of London, Université de Londres
M. Paul Humphreys	Professeur à l'Université de Virginie
M. Thierry Martin	Professeur à l'Université de Franche-Comté
M. Wolfgang Spohn	Professeur à l'Université de Constance

Pré-rapporteurs :

M. Thierry Martin	
M.Glenn Shafer	Professeur à l'Université Rutgers et à Royal Holloway College, Université de Londres

Causalité et probabilités : réseaux bayésiens, propensionnisme

Isabelle Drouet

Thèse
en vue de l'obtention du
Doctorat de l'Université Paris I
en Philosophie

Table des matières

Remerciements

Ce n'est pas seulement par convenance que je commence ici par remercier Jacques Dubucs. Son influence a été décisive à bien des égards. Il m'a donné le goût de la philosophie rigoureuse dans les premières années de mes études à Paris 1, puis m'a orientée vers la philosophie des probabilités à l'occasion de mon DEA. Au cours de mes années de thèse, son soutien, ses conseils et la confiance qu'il m'a accordée ont été précieux. Enfin, sans lui, je n'aurais pas réussi à terminer ce travail dans le temps imparti. De tout cela, donc, je le remercie sincèrement.

Je remercie Glenn Shafer et Thierry Martin d'avoir accepté de rédiger un rapport sur mon travail de thèse, ainsi que Donald Gillies, Paul Humphreys et Wolfgang Spohn de participer à son jury. Mon travail a par ailleurs bénéficié de l'attention et des remarques de Fabien Accominotti, Guido Bacciagaluppi, Donald Gillies et Paul Humphreys à nouveau, Philippe Mongin, Federica Russo, John Vickers ; qu'ils en soient remerciés.

L'IHPST a constitué un environnement particulièrement favorable pour la préparation de ce travail. Le séminaire *Probabilités, décision, incertitude* (anciennement séminaire *Probabilités*) a été pour moi un lieu capital de formation et de discussion. Merci à ses participants, spécialement Mikaël Cozic pour l'*imaging* et pour l'acuité de ses remarques. Merci aussi à Thierry Martin pour sa bienveillance jamais démentie et ses encouragements toujours bienvenus. Au-delà du séminaire *Probabilités, décision, incertitude*, les chercheurs de l'IHPST se sont montré notablement disponibles. Je tiens à remercier en particulier Anouk Barberousse pour son écoute, ses relectures et ses encouragements, Philippe Huneman pour ses conseils en matière de causalité, Jean Mosconi pour l'attention qu'il porte depuis plusieurs années à l'avancement de mes travaux. Je ne saurais non plus manquer de remercier Peggy Tessier-Cardon, qui rend la vie à l'IHPST tellement plus facile et dont la constance et l'amabilité sont un soutien plus grand qu'elle ne se le figure. Enfin, parmi les doctorants et les post-doctorants de l'IHPST, je tiens à remercier en particulier, pour leur compagnie au cours de mes années de thèse ou pour leur aide dans les dernières semaines de rédaction : Adrien Barton,

Jindrich Cerny, Henri Galinon, Élodie Giroux, Neil Kennedy, Marie-Claude Lorne, Francesca Merlin, Cédric Paternotte, Thomas Pradeu, Carlo Proietti, Arancha San Gines Ruiz, Guy-Cédric Werlings.

Pour la préparation de ma thèse j'ai bénéficié d'une allocation de recherche et d'un monitorat attribués par l'Université Paris 1. Je remercie à ce titre l'UFR de philosophie. Je remercie également l'École doctorale de philosophie de Paris 1 et l'IHPST pour avoir financé ma participation à plusieurs formations et conférences.

Merci à mes amis dont la présence m'a permis de mener ce travail à bien. Je pense surtout à Anne, Fabien, Léonie et Sophie. Merci à ma famille, grand-parents de coeur compris, spécialement mes parents pour la confiance qu'ils m'ont témoignée en cette entreprise difficile, et ma soeur pour les attentions dont elle a émaillé sa réalisation. Merci, enfin, à Nicolas, qui m'a soutenue tout au long de mes années de thèse, et sans qui ce travail n'aurait su être le même.

Paris, novembre 2007

Avertissement

Ce travail de thèse comporte deux annexes. Elles sont toutes deux rédigées en langue anglaise et visent principalement à faciliter la lecture de ce travail par ceux des membres du jury qui ne sont pas, ou pas parfaitement, francophones. Leur présence n'enlève rien à l'autonomie de la partie rédigée en français, qui constitue la thèse proprement dite.

L'annexe A se compose de trois articles que j'ai rédigés au cours de ma thèse. Pour chacun de ces articles, son contenu a fait l'objet d'une ou plusieurs présentations, et il est intégré à un chapitre de la thèse.

L'article A.1, "Causal inference : How can Bayes nets contribute" (Drouet (2007)), a été écrit à l'occasion de la conférence "Causality and probability in the sciences" qui s'est tenue à Canterbury en juin 2006. Je traite la même question dans ce texte et dans le chapitre 3 de la thèse. L'argument principal (celui qui est présenté dans le paragraphe 3.3.1.2 de la thèse) est le même dans les deux cas. Toutefois, les analyses du chapitre 3 sont sensiblement plus fouillées et nuancées que celles de l'article. Sous cette forme plus achevée, elles ont été présentées à l'occasion de l'atelier « Causalité » organisé à l'IHPST en octobre 2007 par Mikaël Cozic et Philippe Mongin.

L'article A.2, "Is determinism more favourable than indeterminism for the causal Markov condition ?", n'est pas publié à ce jour. Il a été présenté en avril 2006 lors d'une journée "Philosophy, Probability, Physics" organisée par Guido Bacciagaluppi et Roman Frigg. Le contenu du chapitre 4 coïncide très largement avec celui de l'article. Il en diffère essentiellement par une analyse plus poussée du concept de déterminisme tel qu'il est mobilisé dans les débats relatifs aux réseaux bayésiens.

L'article A.3, "Can there be a propensity interpretation of conditional probabilities ?", a été présenté au séminaire PDI de l'IHPST en janvier 2007, puis à la conférence "Reasoning about probabilities and probabilistic reasoning" qui a eu lieu à Amsterdam en mai 2007. Il n'est pas publié, mais en cours d'évaluation. Son contenu est intégré au chapitre 6. Le chapitre, toutefois, est plus riche que l'article. D'un côté l'analyse du paradoxe de Humphreys et de l'autre celle des conséquences de la proposition que nous

formulons (en particulier celles qui sont relatives au rapport entre la causalité et les probabilités) sont explorées beaucoup plus longuement dans le chapitre que dans l'article.

L'annexe B est un long résumé de la thèse. Plus précisément, ce résumé commence par la traduction extensive de l'introduction à la thèse et continue avec un résumé de chacun des chapitres puis de la conclusion. Le plan du résumé correspond exactement à celui de la thèse ; il est seulement moins détaillé.

Les références des textes mentionnés dans les annexes ont été reportées à la bibliographie générale. Le lecteur les retrouvera aisément même dans le cas où elles ne sont pas abrégées de la même façon dans les annexes et dans la bibliographie de la thèse. L'index des noms et l'index des notions ne tiennent pas compte des annexes. Enfin, j'ai traduit les citations anglaises du corps du texte. Pour les mots ou expressions qui n'ont pas d'équivalent français immédiat, j'ai indiqué le mot anglais ou l'expression anglaise entre parenthèses à la suite de la première occurrence de la traduction que j'en propose.

Causalité et probabilités : réseaux bayésiens, propensionnisme

Introduction

Il n'est pas douteux que la causalité joue un rôle central dans nos explications, qu'elles soient scientifiques ou non. Il n'est pas douteux non plus que l'efficacité de nos actions dépend de notre connaissance des causes. En revanche, il est simplement faux que nous en ayons fini avec l'analyse de cette notion. Cela ne signifie pas que la causalité n'ait pas intéressé les philosophes ; au contraire, elle a été un objet d'intérêt tout au long de l'histoire de la philosophie. Cela signifie plutôt que nous disposons d'un nombre considérable de tentatives d'explicitation de ce qu'est la causalité, que la façon dont ces tentatives s'articulent n'est pas toujours claire, et surtout qu'aucune d'entre elles n'a réussi à former autour d'elle un consensus.

S'il est difficile de s'orienter au sein du champ des tentatives d'élucidation de ce qu'est la causalité, il est néanmoins possible d'identifier un élément saillant de son développement récent. Pour le comprendre, commençons par noter que l'idée selon laquelle la causalité serait une relation de nécessitation est restée pratiquement non discutée jusqu'au milieu du vingtième siècle. Anscombe (1981) explique que « la vérité de cette conception n'est guère débattue. Elle est, en fait, un morceau de *Weltanschauung* »[1]. Plus précisément, le début de l'article montre comment la thèse selon laquelle la causalité est une relation de nécessitation a traversé l'histoire de la philosophie occidentale, d'Aristote à Russell. Elle explique en particulier que la critique humienne de l'idée selon laquelle la causalité serait une relation *logique* n'a pas défait l'attachement de la causalité à la nécessitation : « en ce qui concerne l'identification de la causalité avec la nécessitation, la pensée de Hume ne l'a pas affaiblie mais, curieusement, l'a renforcée »[2].

Or, précisément, cette identification a été remise en cause dans les années 1960. Plus explicitement, des théories *probabilistes* de la causalité sont apparues dans les années 1960, et se sont développées à partir des deux thèses suivantes : d'une part certaines causes ne rendent pas leurs effets nécessaires, mais d'autre part les causes doivent pouvoir être caractérisées par ceci qu'elles

[1] Anscombe (1981) p. 89.
[2] Anscombe (1981) p. 89–90.

augmentent la probabilité de leurs effets. Ce sont ces théories qui nous intéressent dans le travail qui commence. Plus précisément, les questions que nous y traitons se posent dans le domaine théorique qui apparaît à l'occasion de la formulation de théories probabilistes de la causalité.

La prochaine section présente ces théories. Plus exactement, elle en présente ce qui motive notre travail. Cette présentation fait apparaître que la distinction entre deux types de causes, génériques et singulières, est centrale dans le cadre des théories probabilistes de la causalité. Aussi la distinction entre causalité générique et causalité singulière est-elle discutée dans une deuxième section. Les troisième et quatrième sections sont consacrées à exposer les problèmes que nous traiterons. Les uns sont relatifs à la causalité générique, les autres à la causalité singulière.

Théories probabilistes de la causalité

L'introduction de concepts probabilistes dans l'analyse de la causalité rompt l'attachement séculaire de la causalité à la nécessitation. Plus précisément, les théories probabilistes de la causalité s'inscrivent dans la tradition d'analyse de la causalité qui trouve son point d'origine chez Hume, et c'est au sein de cette tradition qu'elles défont le lien qui attache la causalité à la nécessitation. Pour Hume, la causalité existe d'abord sous la forme de régularités et se caractérise en particulier par la conjonction constante[3] : une cause est invariablement suivie de son effet. Cette thèse, clairement, n'est pas satisfaisante : elle implique que gratter une allumette ne cause pas l'embrasement de l'allumette dès lors qu'il existe des situations dans lesquelles gratter une allumette n'est pas suivi de l'embrasement de l'allumette. Plus généralement, la thèse humienne selon laquelle une cause est invariablement suivie de son effet est incapable de rendre compte de ceci : la plupart des causes ne suffisent pas à produire leurs effets, mais ne les produisent qu'en présence de certains facteurs (qui sont également appelées « causes » sous la plupart des analyses de la causalité). Dans le cas de l'allumette, un de ces facteurs est la présence d'oxygène dans le milieu dans lequel elle est grattée.

Le problème que nous venons de mettre au jour ne requiert pas, pour être traité, l'introduction de probabilités dans l'analyse de la causalité. Ainsi, il est traité dans le cadre d'analyses régularistes du concept de cause plus raffinées que celle qui est proposée par Hume. La plus connue de ces analyses est celle de Mackie : une cause est une « partie *insuffisante* mais *non redondante* d'une condition *non nécessaire* mais *suffisante* »[4] pour son effet, une « condition » étant ici un ensemble de facteurs.

[3] Hume (1739) p. 150.

[4] Mackie (1974) p. 62. Les italiques sont dans le texte original.

Raffinées ou non, les analyses régularistes supposent qu'il n'y a pas d'effet sans un ensemble de facteurs qui suffit à le produire – et donc qu'il suit régulièrement. Corrélativement, il n'y a de cause qu'appartenant à un ensemble de facteurs qui suffit à produire son effet. Or, il n'est pas clair que toutes les causes sont bien de ce type, qu'il n'existe pas de causes qui produisent leur effet sans appartenir à un ensemble de facteurs suffisant à cette production. Plus précisément, on considère généralement que l'hypothèse selon laquelle de telles causes existent a pris consistance avec la découverte des phénomènes quantiques et qu'elle est aujourd'hui très plausible, y compris hors du domaine quantique. Ainsi, nous ne connaissons pas d'ensemble de facteurs auquel appartiendrait la propriété d'être fumeur et qui suffirait à produire le cancer du poumon. Surtout, rien ne garantit qu'un tel ensemble existe. De façon générale, l'hypothèse selon laquelle il existe des effets sans ensemble de facteurs qui suffit à les produire est au moins très plausible, sinon avérée.

Contrairement aux analyses régularistes, les théories probabilistes de la causalité sont compatibles avec cette hypothèse. En effet, elles reposent sur la proposition de caractériser une cause par ceci qu'elle rend son effet plus probable. Plus précisément, une cause C augmente la probabilité de son effet E au sens où la probabilité conditionnelle $p(E|C)$ est plus élevée que la probabilité absolue $p(E)$. Nous verrons plus loin[5] que cette idée doit être raffinée pour fonder effectivement une analyse de la causalité. Nous nous contentons ici de souligner, d'abord, ceci : dire qu'une cause rend son effet plus probable n'implique ni qu'elle lui donne une probabilité de 1, ni qu'elle appartient à un ensemble de facteurs qui lui donne cette probabilité. L'idée qu'on trouve au fondement des théories probabilistes de la causalité est donc bien telle qu'il peut exister des effets sans ensemble de facteurs qui suffit à les produire.

Au premier rang des questions soulevées par les théories probabilistes de la causalité, on trouve celle de leur statut. Sur ce point, nous défendons trois thèses qui ne sont pas indépendantes. D'abord, les théories probabilistes de la causalité sont des *analyses conceptuelles*. Il s'agit de déterminer ce que cela veut dire que A cause B.[6] En particulier il ne s'agit pas – ou alors seulement dans un second temps – de donner des critères permettant de reconnaître pratiquement les relations de causalité. Il ne s'agit pas non plus de définir la causalité. En effet, la seconde des thèses que nous défendons relativement au statut des théories probabilistes de la causalité est la suivante : elles sont

[5] Dans la section 2.2.

[6] Cette définition de l'analyse conceptuelle pourrait être discutée (voir Humphreys (1989) p. 3 pour une définition différente), mais elle nous semble suffisamment claire pour que nous l'utilisions ici.

des analyses *d'un aspect* du concept de causalité. Pour le dire autrement, les théories probabilistes de la causalité répondent à *une* question relative à la causalité. On pourrait formuler cette question dans les termes suivants : quel est le rapport de co-occurrence des causes et de leurs effets ? Cette question diffère en particulier de la question de savoir à quel type de réalités appartiennent les causes, les effets et les relations de cause à effet. Enfin, les théories probabilistes de la causalité telles que nous venons de les caractériser se présentent en premier lieu comme des théories de la causalité *générique*, plutôt que comme des théories de la causalité singulière.

Les questions que nous traitons dans ce travail dépendent très largement de cette dernière remarque. En effet, la causalité générique et la causalité singulière ne sont pas dans des situations similaires relativement aux théories probabilistes. Nous le montrons dans la prochaine section.

Causalité générique et causalité singulière

Distinction de la causalité générique et de la causalité singulière. Du point de vue des énoncés, la différence entre causalité générique et causalité singulière est la différence qui existe entre « Fumer cause le cancer du poumon » et « Le tabagisme de Pierre a causé son cancer du poumon », ou entre « Les chutes causent des fractures » et « Ma chute dans l'escalier ce matin a causé la fracture de mon poignet droit ». Ainsi, la causalité générique est une relation entre propriétés – par exemple la propriété de chuter et la propriété de souffrir d'une fracture. De l'autre côté, la causalité singulière est une relation entre événements – singuliers – qui sont effectivement advenus. Causalité générique et causalité singulière sont généralement désignées comme des *niveaux* (*levels*) de causalité ; par commodité, nous nous conformerons à cet usage.

La question du rapport entre niveaux de causalité reste débattue. Les trois réponses principales consistent à considérer pour la première que les relations de cause à effet singulières ne sont causales que parce qu'elles instancient des relations de cause à effet génériques, pour la deuxième que la causalité générique tire sa réalité de la causalité singulière[7], pour la troisième que la causalité générique et la causalité singulière sont des réalités indépendantes. Nous ne prendrons pas parti en faveur de l'une ou l'autre de ces réponses. Plus généralement, la question du rapport entre causalité générique et causalité singulière n'est pas abordée pour elle-même dans ce travail.

En effet, la réponse qu'on apporte à cette question est neutre relativement à ce qu'il y a à dire des théories probabilistes de la causalité, et en particulier

[7]Une version de cette réponse consiste à considérer que les affirmations causales génériques sont des généralisations d'affirmations causales singulières.

de leur opportunité et de leur pertinence. A titre de justification de cette affirmation, nous remarquons que chacune des trois réponses principales à la question du rapport entre niveaux de causalité a été défendue par des tenants de théories probabilistes de la causalité. La première réponse est impliquée par la prétention de Suppes à étendre l'analyse de Hume aux causes qui ne suffisent pas à produire leurs effets[8]. En effet, chez Hume les relations de cause à effet singulières n'existent *comme relations causales* que pour autant qu'elles instancient les relations de cause à effet génériques.[9] La deuxième réponse est à l'oeuvre dans Cartwright (1989) et dans Humphreys (1989). La troisième réponse est défendue en particulier dans Eells (1991) et dans Sober (1985), avec des conséquences différentes : Sober réserve l'analyse probabiliste à la causalité générique alors que Eells développe deux théories probabilistes, l'une pour la causalité singulière et l'autre pour la causalité générique.

Nous venons d'expliquer pourquoi s'intéresser aux théories probabilistes de la causalité n'impose pas de prendre position sur la question du rapport entre niveaux de causalité. Ce qu'en revanche nous ne pouvons éviter est de formuler la remarque suivante : la causalité générique et la causalité singulière ne sont pas dans des situations similaires relativement aux théories probabilistes. Pour le dire autrement, il existe plusieurs raisons pour lesquelles il n'existe pas d'analogie entre des théories probabilistes de la causalité générique d'un côté et des théories probabilistes de la causalité singulière de l'autre. La fin de la présente section vise à expliciter cette thèse en même temps qu'à présenter des théories probabilistes de la causalité générique et des théories probabilistes de la causalité singulière, ce qui nous est nécessaire pour faire apparaître les questions que nous traitons dans le travail qui commence.

Causalité générique et théories probabilistes. En vue de comprendre que causalité générique et causalité singulière ne sont pas dans des situations similaires relativement à l'analyse probabiliste, commençons par revenir sur la thèse énoncée plus haut, selon laquelle les théories probabilistes se présentent en premier lieu comme des théories de la causalité générique. La première précision qu'il convient que nous apportions alors est la suivante : l'emploi que nous faisons de l'expression « en premier lieu » n'est pas temporel. Ainsi, l'analyse proposée par Suppes dans l'ouvrage véritablement fondateur du champ des théories probabilistes de la causalité est conçue par son auteur comme valant aussi bien de la causalité générique que de la cau-

[8]Suppes (1970) p. 9.

[9]Voir Hume (1739) p. 150.

salité singulière.[10] De l'autre côté, c'est seulement au tournant des années 1990 que les théories probabilistes sont présentées explicitement soit comme des théories de la causalité générique, soit comme des théories de la causalité singulière.[11]

Venons-en maintenant à ce que nous visions, positivement, en soutenant que les théories probabilistes de la causalité se présentent en premier lieu comme des théories de la causalité générique. Il s'agit principalement de ceci que les contre-exemples à l'idée de caractériser une cause par ce qu'elle augmente la probabilité de son effet concernent la causalité singulière. Ainsi du contre-exemple le plus fameux, introduit par Rosen[12] : Jones, un joueur de golf médiocre, frappe la balle, laquelle heurte une branche d'un arbre environnant, mais la heurte de telle sorte que la balle tombe dans le trou. Le tir de Jones *cause* bien la chute de la balle dans le trou alors même qu'elle en diminue la probabilité. Par ailleurs, cette relation de cause à effet est spécifiquement singulière – et avec elle le problème qu'elle constitue pour les théories probabilistes de la causalité : c'est *ce* tir de Jones qui cause la chute de la balle dans le trou, et non les tirs du médiocre Jones de manière générale.

De manière plus générale, nous soutenons que les théories probabilistes de la causalité générique trouvent des formulations satisfaisantes dans Cartwright (1989)[13] et Eells (1991)[14]. D'une part, en effet, ces ouvrages développent des théories qui rendent compte de manière satisfaisante de tous les cas identifiés comme problématiques pour les théories probabilistes de la causalité antérieures. D'autre part, et de façon corrélative, le débat portant sur les théories probabilistes de la causalité générique s'est tari à la suite de la publication de ces deux analyses. Selon cette thèse, nous savons ce que veut dire « fumer cause le cancer du poumon ».

Causalité singulière et théorie probabiliste. Du côté de la causalité singulière, les choses se présentent différemment. Nous venons de voir que la causalité singulière se prête moins bien que la causalité générique à une analyse fondée sur la notion d'augmentation de probabilités conditionnelles. Plus précisément, nous avons vu qu'un tir singulier de Jones peut causer la chute de la balle de golf dans le trou alors même qu'il a diminué la probabilité de cet événement. Toutefois il existe au moins une théorie probabiliste solide

[10] Suppes (1970) p. 75.

[11] En particulier, Cartwright (1989), Humphreys (1989) et Eells (1991).

[12] L'exemple est déjà mentionné dans Suppes (1970) (p. 41), mais Rosen ne le rend public que dans Rosen (1978) (pp. 607–608).

[13] Cartwright (1989) chap. 3 et 4.

[14] Eells (1991) partie 1.

de la causalité singulière : celle qui est développée dans Humphreys (1989)[15]. Or, cette théorie présente trois caractéristiques remarquables, qui la rendent critiquable et impliquent que la causalité singulière n'est pas, relativement à l'analyse probabiliste, dans une situation similaire à celle de la causalité générique.

En premier lieu, la théorie de Humphreys conduit parfois à des verdicts causaux qui ne sont pas intuitifs. Ce point est souligné en particulier par Woodward[16]. Humphreys répond par avance qu'en cas de conflit entre la théorie philosophique systématique et le langage ordinaire, la première doit prendre le pas sur le second.[17] Mais cet argument n'a pas suffi à convaincre qu'Humphreys (1989) rend complètement tenable l'idée de caractériser une cause singulière par l'augmentation de la probabilité de son effet. Plus généralement, les arguments développés dans Sober et Eells (1983) et dans Sober (1985) en faveur de l'idée selon laquelle la seule causalité générique est susceptible d'une analyse probabiliste continuent d'être largement considérés comme corrects.

En deuxième lieu, la théorie de Humphreys est une théorie de la causalité entre instanciations de propriétés. Autrement dit, elle conduit à considérer qu'un événement susceptible d'être cause ou effet est toujours l'instanciation d'une propriété par un système : « un *événement* est un changement dans, ou la possession de, une propriété par un système lors d'un essai (*trial*) »[18]. Les événements singuliers sont donc définis en référence à ce qu'ils ne sont pas, et plus précisément en référence aux *relata* de la causalité générique. A cela, on peut opposer l'idée selon laquelle les événements singuliers, pour être vraiment singuliers, doivent être définis par ce qu'ils sont spécifiquement : effectivement actualisés dans l'espace et dans le temps physiques. Cette exigence semble d'autant plus légitime qu'elle est compatible avec la caractérisation des événements que propose Quine. Cette caractérisation, qui est devenue classique, stipule que : « chaque événement inclut le contenu, aussi hétérogène qu'il soit, de quelque portion de l'espace-temps, aussi déconnectée et étrange qu'elle soit »[19]. La théorie de Humphreys laisse au moins ouverte la question de savoir comment elle s'applique à des événements singuliers ainsi conçus.

En troisième lieu, les probabilités mobilisées par Humphreys pour l'analyse de la causalité singulière sont des probabilités physiques.[20] Le recours à des probabilités physiques est justifié par l'objet de l'analyse : il s'agit de

[15]Humphreys (1989) § 31.
[16]Woodward (1994) pp. 366–367.
[17]Humphreys (1989) p. 5.
[18]Humphreys (1989) p. 24.
[19]Quine (1960) p.171.
[20]Humphreys (1989) p. 54 en particulier.

la causalité probabiliste en tant que caractéristique du monde.[21] Il a pour conséquence que les probabilités que Humphreys utilise pour analyser la causalité singulière reçoivent une interprétation propensionniste. En effet, parmi les interprétations des probabilités, seule l'interprétation propensionniste permet de penser des probabilités physiques d'événements singuliers. Mais si cette interprétation est la seule qui nous permet de penser des probabilités physiques d'événements singuliers, elle est également la seule dont nous avons de bonnes raisons de douter qu'elle permet de penser les probabilités *conditionnelles*.[22] Or nous avons vu que la notion d'augmentation de probabilités qui fonde les théories probabilistes de la causalité se comprend en termes de probabilités conditionnelles. Nous avons donc mis au jour un troisième point sur lequel la théorie de Humphreys peut être discutée.

Ainsi que nous l'avons annoncé, les questions traitées dans notre travail sont des questions qui se posent dans l'état actuel de développement des théories probabilistes de la causalité. Or, nous venons de montrer que, relativement à ces théories, la causalité générique et la causalité singulière ne sont pas dans des situations similaires. Il en découle que les théories probabilistes soulèvent des questions très différentes selon qu'on s'intéresse à la causalité générique ou à la causalité singulière. En conséquence, notre travail se compose de deux parties : la première traite de questions ouvertes par les théories probabilistes de la causalité générique, la seconde porte sur des questions relatives à l'analyse probabiliste de la causalité singulière.

Théories probabilistes et épistémologie de la causalité générique

S'il est vrai que nous disposons, avec les théories probabilistes, d'analyses satisfaisantes du concept de causalité générique, alors il semble naturel d'en venir au problème des critères de reconnaissance des causes. C'est en tout cas ce que suggère la distinction, introduite plus haut, entre trois choses : analyser le concept de cause, donner un critère de reconnaissance de la causalité et définir la causalité. Dans le cadre de cette distinction, nous avons soutenu que les théories probabilistes sont des analyses conceptuelles et qu'elles visent seulement un aspect du concept de cause. Dans ces conditions, il ne saurait y avoir de définition probabiliste de la causalité générique. La seule question que laisse ouverte l'achèvement de l'analyse du concept de causalité générique est donc la question épistémologique des critères de reconnaissance des relations de cause à effet.

[21] Humphreys (1989) pp. 54–55.

[22] Humphreys (1985), Milne (1986), Humphreys (2004).

Cette question ne se pose pas seulement de manière négative et ne se justifie pas seulement par l'existence d'un problème non traité. Positivement, elle se justifie aussi par les relations qu'entretiennent l'analyse conceptuelle et la formulation de critères de reconnaissance : disposer d'une analyse de « A cause génériquement B » semble un bon point de départ en vue de la formulation de critères de reconnaissance des causes génériques. La question, alors, se précise : les théories probabilistes de la causalité donnent-elles des critères utilisables de reconnaissance des causes génériques et, si oui, lesquels ? Pour le dire en termes méthodologiques : existe-t-il des méthodes d'inférence aux causes génériques qui mobilisent des critères de reconnaissance hérités des théories probabilistes de la causalité ?

Il existe un candidat à ce titre qui est naturel, en même temps que digne d'attention : les méthodes d'inférence causale fondées sur la notion de réseau bayésien. Ces méthodes, issues de l'intelligence artificielle, apparaissent au tournant des années 1990. Elles consistent essentiellement en des algorithmes qui construisent un graphe à partir d'informations probabilistes du type de celles que donnent les statistiques. Ce graphe est orienté et les flèches qui y figurent sont interprétées causalement, c'est-à-dire qu'on considère que chacune représente une relation de cause à effet générique. De cette première description très abstraite, il découle que les méthodes d'inférence auxquelles nous faisons allusion doivent mobiliser un critère probabiliste de causalité. Ce critère doit pouvoir être confronté aux théories probabilistes de la causalité. L'idée d'une telle confrontation apparaît d'autant plus pertinente que ce critère entretient une nette parenté avec les théories probabilistes de la causalité. Ainsi il n'est pas rare que des textes de présentation du champ des analyses de la causalité fassent entrer ledit critère au nombre des théories probabilistes de la causalité. C'est le cas par exemple de Hitchcock (2002).

La question se pose alors du rapport exact entre d'une part le critère de causalité qu'utilisent les méthodes d'inférence causale fondées sur les réseaux bayésiens et d'autre part les théories probabilistes de la causalité. Cette question se pose d'autant plus naturellement que le critère de causalité véhiculé par les réseaux bayésiens s'identifie aisément. Ainsi Cartwright reconnaît-elle qu'il s'agit de celles « des méthodes d'inférence causale disponibles [qui sont] fondées le plus explicitement et avec le plus de soin »[23]. C'est que les réseaux bayésiens, sur lesquels ces méthodes sont fondées, se définissent précisément par le rapport entre les probabilités et les graphes interprétés causalement. Ce rapport est discuté dans le chapitre 1. Plus généralement, ce premier chapitre est consacré à présenter les réseaux bayésiens et à discuter leur interprétation causale.

[23]Cartwright (1999) p. 20.

Revenons-en à la question plus spécifique du rapport entre les théories probabilistes de la causalité et l'inférence aux causes telle qu'elle est autorisée par les réseaux bayésiens. Cette question se pose de façon d'autant plus pressante que les théories probabilistes de la causalité ne donnent pas lieu immédiatement à des critères utilisables pour identifier les causes. En effet, dans leur forme aboutie, elles sont des analyses complexes et circulaires – au sens où l'analyse du concept de cause mobilise, on le verra[24], des concepts causaux. Dès lors, on se demande pourquoi et comment les réseaux bayésiens permettent d'inférer des causes. Est-ce ceci que le critère de reconnaissance de la causalité n'est pas, finalement, apparenté aux théories probabilistes ? Ou alors la circularité et la complexité n'empêchent-elles pas les théories probabilistes de la causalité générique de donner un critère de reconnaissance des causes qui soit utilisable ? Dans ce dernier cas, quel rôle la composante graphique des réseaux bayésiens joue-t-elle relativement à la possibilité de mener effectivement l'inférence aux causes ? Ces questions sont abordées dans le chapitre 2.

Du fait précisément de leur composante graphique, les réseaux bayésiens causaux s'inscrivent dans une tradition inaugurée par Wright dans les années 1920. Cette tradition est celle des modèles causaux utilisés à des fins d'épistémologie de la causalité générique. Les partisans des réseaux bayésiens soutiennent que l'originalité des réseaux bayésiens considérés comme modèles causaux est qu'ils permettent d'*induire* des causes. Si tel est bien le cas, alors les réseaux bayésiens renouvellent considérablement l'épistémologie de la causalité générique, qui est historiquement attachée à l'hypothético-déduction. Qu'en est-il exactement ? Les méthodes d'inférence aux causes fondées sur les réseaux bayésiens sont-elles effectivement inductives ? Plus généralement, qu'apportent les réseaux bayésiens à la méthodologie de l'inférence causale générique et en quoi les méthodes d'inférence causale fondées sur les réseaux bayésiens diffèrent-elles de méthodes plus traditionnelles qu'elles peuvent concurrencer ? Ces questions sont traitées dans le chapitre 3.

Les questions que nous abordons dans le chapitre 3 prennent une importance accrue quand on les conçoit comme des sous-questions d'une question plus générale, celle des modalités de l'inférence aux causes à partir de données probabilistes. Cette question a une histoire qui remonte au moins à la fin du dix-neuvième siècle et à la constitution de la sociologie française. En effet, la sociologie s'est émancipée de la psychologie à partir de réflexions méthodologiques, portant pour certaines sur la possibilité d'utiliser des données statistiques pour tirer des conclusions causales – deux points

[24] Dans la section 2.2.

dont Durkheim (1895) témoigne.[25] Le chapitre 3 revient donc sur des questions classiques et les traite à la lumière nouvelle des réseaux bayésiens.

De façon sensiblement différente, le chapitre 4 s'inscrit de plain-pied dans les débats les plus contemporains portant spécifiquement sur les réseaux bayésiens. Au fond, la question est de savoir quand sont satisfaites les hypothèses qui garantissent que les méthodes d'inférence causale fondées sur les réseaux bayésiens peuvent être utilisées. Plus précisément, il s'agit de caractériser les systèmes réels qui satisfont ces hypothèses. Concernant la principale de ces hypothèses, appelée « condition de Markov causale », il a semblé d'abord qu'elle était plus susceptible d'être satisfaite par les systèmes déterministes que par les systèmes indéterministes. Cette thèse a été soutenue par ceux-là même qui ont introduit et défendent les réseaux bayésiens causaux. Elle n'a pas semblé problématique jusqu'à Steel (2005), qui la remet en cause. Le déterminisme est-il ou n'est-il pas plus favorable que l'indéterminisme pour la condition de Markov causale ? Nous nous attachons à le décider dans le chapitre 4, après quoi nous en venons à la seconde partie de notre travail. Nous présentons maintenant les problèmes qui y sont traités.

Théories probabilistes de la causalité singulière

Nous avons expliqué plus haut que la causalité générique et la causalité singulière ne sont pas dans des situations similaires relativement à l'analyse probabiliste. Plus précisément, nous avons indiqué trois points sur lesquels notre meilleure théorie probabiliste de la causalité singulière – celle qui est développée dans Humphreys (1989) – peut encore être discutée. Dans ces conditions, l'état du développement des théories probabilistes de la causalité impose dans le cas singulier des questions différentes de celles qui sont apparues pour le cas générique. Plus explicitement, la seconde partie ne traite pas de questions épistémologiques qui seraient analogues à celles qui nous occupent dans la première partie. Positivement, nous en restons au plan de l'analyse conceptuelle, c'est-à-dire à la discussion de ce que cela veut dire que A cause singulièrement B, que le tir de Jones a causé la chute de la balle dans le trou ou que le tabagisme de Pierre a causé son cancer du poumon.

Nous abordons cette discussion à partir de deux des trois remarques que nous avons formulées plus haut concernant Humphreys (1989). La première de ces remarques concerne la notion d'événement, et plus précisément l'opportunité de définir les événements en référence aux propriétés. Nous avons fait valoir que caractériser un événement comme le contenu d'une zone spatio-temporelle fait mieux droit à ce que les événements singuliers ont

[25] Voir en particulier le dernier chapitre.

de spécifique, et qui du coup contribue à distinguer la causalité singulière de la causalité générique. Le projet qui se fait alors jour est celui d'une analyse probabiliste de ce que nous appellerons « causalité actuelle », c'est-à-dire de la causalité singulière considérée comme relation entre des événements définis par ce qui fait leur singularité. En amont de la production d'une telle analyse, la question qui se pose est celle du rapport entre la causalité actuelle et les probabilités. La seconde partie de notre travail vise cette question et lui apporte des éléments de réponse.

Nous abordons la question du rapport entre causalité actuelle et probabilités à partir de la philosophie des probabilités. Cette approche se justifie par ce qu'elle permet de déterminer la question encore vague du rapport entre causalité actuelle et probabilités. Plus précisément, nous soutenons que les probabilités n'ont de rapport avec la causalité actuelle – et au-delà ne peuvent en fonder une analyse – que si elles reçoivent une interprétation propensionniste. En effet, c'est la seule parmi les interprétations des probabilités disponibles qui permet de penser des probabilités d'événements singuliers qui sont des caractéristiques du monde physique au titre où l'est la causalité actuelle. Dans ces conditions, il est vrai qu'on détermine la question du rapport entre causalité actuelle et probabilités si on l'aborde à partir de la philosophie des probabilités. Plus explicitement, la question devient celle du rapport entre la causalité et les probabilités interprétées comme des propensions.

La question à laquelle nous aboutissons a un sens pour elle-même, et indépendamment de la question du rapport entre la causalité actuelle et les probabilités. En effet, l'interprétation propensionniste des probabilités telle que l'introduit Popper au milieu des années 1950 repose sur l'idée selon laquelle les probabilités mesurent des dispositions physiques à produire les événements singuliers – les « propensions ». Or sous la conception réaliste à laquelle Popper tend souvent, les propensions ressemblent étrangement à des causes : il s'agit d'entités physiques capables de produire des événements. Les propensions sont-elles des causes et en quel sens ? Qu'en découle-t-il pour l'épistémologie, l'ontologie et même la métaphysique ? Ces questions sont abordées dans le chapitre 5. Ce chapitre a été conçu comme une présentation approfondie du propensionnisme de tradition popperienne, qui prend en compte la question des corrélats philosophiques de cette théorie. En conséquence, nous abordons dans ce chapitre des problématiques de philosophie plus générale que celles qui sont traditionnellement attachées à la seule philosophie des probabilités.

Le chapitre 5 s'en tient au propensionnisme considéré comme théorie des probabilités *absolues*. Or, nous avons vu plus haut que l'explicitation de l'idée séminale pour les théories probabilistes de la causalité, selon laquelle une cause se caractérise par ce qu'elle augmente la probabilité de son effet, fait

apparaître des probabilités *conditionnelles*. Il en découle que la seconde partie de notre travail aura le sens nous lui avons assigné seulement si nous ne nous en tenons pas au cas absolu, et si nous faisons porter l'analyse sur le rapport entre la causalité et les probabilités conditionnelles quand elles reçoivent une interprétation propensionniste.

C'est ici que nous rencontrons la troisième des remarques que nous formulions plus haut relativement à la théorie de Humphreys. Cette remarque consistait plus précisément à faire valoir que Humphreys recourt à la fois à des probabilités conditionnelles et à des propensions, là où il existe des objections solides à l'idée d'interprétation propensionniste des probabilités conditionnelles. Nous y avons vu plus haut le fondement d'une critique possible de la théorie de Humphreys. Nous y voyons maintenant une menace pour notre projet d'extension des analyses du chapitre 5 au cas des probabilités conditionnelles. Cette menace sera d'autant plus sérieuse que le chapitre 5 aura plus mis l'accent sur la dimension causale de l'interprétation propensionniste popperienne des probabilités absolues, et que c'est de cette dimension causale que découlent les difficultés du propensionnisme avec les probabilités conditionnelles. Pour le dire en termes comparatifs : le propensionnisme que nous aurons présenté dans le chapitre 5 est beaucoup plus sensible que le propensionnisme de Humphreys (1989) aux difficultés soulevées par les probabilités conditionnelles.

Du coup il semble que nous devons abandonner le projet de penser le rapport entre causalité actuelle et probabilités conditionnelles. Plus précisément, nous devons abandonner ce projet sauf s'il est possible de contourner l'obstacle que les probabilités conditionnelles constituent pour le propensionnisme. Le chapitre 6 est très largement consacré à explorer les voies d'un contournement possible de cet obstacle. Nous y discutons les raisons de penser que le propensionnisme ne permet pas d'interpréter les probabilités conditionnelles puis proposons ce qui pourrait être une interprétation propensionniste des probabilités conditionnelles. Que découle-t-il de cette proposition relativement au rapport entre causalité actuelle et probabilités conditionnelles ? Comment cela s'articule-t-il avec les analyses du chapitre 5, relatives au rapport entre causalité et probabilités absolues interprétées en termes de propensions ? Ces questions sont abordées à la fin du chapitre 6.

Première partie

Causalité générique et réseaux bayésiens

Chapitre 1

Réseaux bayésiens causaux

Dans le chapitre qui commence, nous présentons les réseaux bayésiens causaux. Cette présentation s'inscrit en tête d'une partie qui vise en particulier à évaluer les méthodes d'inférence aux causes fondées sur les réseaux bayésiens. Il convient donc de mettre l'accent, dans ce premier chapitre, sur les aspects par lesquels la notion même de réseau bayésien causal limite la possibilité d'inférer des causes grâce aux réseaux bayésiens. Ainsi, nous veillerons à faire apparaître ce que la notion de réseau bayésien causal suppose relativement à la causalité en général, et à ses rapports avec les probabilités en particulier. Nous procédons en trois temps :

1. dans la première section, nous présentons les réseaux bayésiens, indépendamment de leur interprétation causale ;
2. dans la deuxième section, nous en venons aux réseaux bayésiens causaux. Nous analysons les hypothèses corrélatives de la notion de réseau bayésien causal et montrons qu'elles sont plausibles ;
3. dans la troisième section, nous envisageons et discutons des contre-exemples qui ont été opposés aux hypothèses corrélatives de la notion de réseau bayésien causal.

Une quatrième et dernière section est consacrée à une présentation synthétique des résultats obtenus.

1.1 Présentation des réseaux bayésiens

1.1.1 D'où viennent les réseaux bayésiens ?

Les réseaux bayésiens sont apparus dans la première moitié des années 1980 et en intelligence artificielle. Plus précisément, ils ont été introduits dans

ce champ comme des outils de traitement de l'incertitude. Dès lors, comprendre d'où viennent les réseaux bayésiens implique de faire retour d'abord sur la question du traitement de l'incertitude en intelligence artificielle.

1.1.1.1 Traitements de l'incertitude en intelligence artificielle

La question du traitement de l'incertitude a émergé en intelligence artificielle dans les années 1970, corrélativement du projet de création de systèmes experts. Un système expert est un programme informatique qui reproduit les mécanismes cognitifs d'experts d'un domaine particulier – et pourraient donc servir d'aide à la décision dans ce domaine. Le projet de création de systèmes experts soulève la question du traitement de l'incertitude dans la mesure où les raisonnements humains en général et les raisonnements d'experts en particulier sont des raisonnements en situation d'incertitude, relativement à des énoncés susceptibles d'exceptions, selon des règles défaisables...

L'un des premiers systèmes experts est le système MYCIN pour le diagnostic des infections sanguines, développé à Stanford par Shortliffe, Buchanan et d'autres au début des années 1970. MYCIN repose sur une base de données (*knowledge base*) constituée de règles de la forme : Si le patient présente tel et tel symptômes, alors une conclusion raisonnable est telle ou telle. A titre d'illustration, Buchanan et Shortliffe indiquent la règle suivante[1] :

IF :	The stain of the organism is gram positive, and The morphology of the organism is coccus, and The growth conformation of the organism is chains
THEN :	There is suggestive evidence (.7) that the identity of the organism is streptococcus.

Il apparaît alors que l'incertitude est prise en compte sous la forme quantitative de l'attribution d'un degré à la conclusion autorisée par la règle. Ce degré est déterminé empiriquement, à partir de la consultation d'experts. Il prend ses valeurs dans l'intervalle $[0\,;\,1]$ mais n'est pas la probabilité conditionnelle du conséquent de la règle relativement à son antécédent :

> interroger l'expert révèle graduellement que malgré son apparente similarité avec une affirmation concernant une probabilité conditionnelle, le nombre 0.7 diffère significativement d'une probabilité. L'expert peut bien accorder que $P(h_1|s_1, s_2, s_3) = 0.7$, mais il devient mal à l'aise quand il essaie d'en tirer la conclusion logique que, du coup, $P(\neg h_1|s_1, s_2, s_3) = 0.3$. Il affirme que les trois observations plaident (au degré 0.7) *en faveur* de la conclusion que l'organisme est un *Streptococcus* et ne devraient pas être considérées comme plaidant (au degré 0.3) *contre* la conclusion que c'est un *Streptococcus*. [...]

[1] Shortliffe et Buchanan (1984) p. 238.

> Il est tentant de conclure que l'expert est irrationnel puisqu'il ne veut pas accepter les implications de ses affirmations probabilistes à leurs conclusions logiques. Une autre interprétation, cependant, est que les nombres qu'il a donnés ne doivent pas être considérés comme des probabilités du tout, qu'ils sont des mesures de jugement qui reflètent un degré de *croyance*.[2]

Les auteurs adoptent cette dernière interprétation et construisent le concept dyadique de confirmation non probabiliste dont ils ont besoin. Ils parlent de « facteurs de certitude ».

1.1.1.2 Caractéristiques du traitement de l'incertitude au moyen des réseaux bayésiens

Le traitement de l'incertitude qu'autorisent les réseaux bayésiens est similaire à celui qui est à l'oeuvre dans les systèmes experts du type de MYCIN sur le point suivant : il n'est pas *logique*, mais repose sur l'introduction d'un concept *numérique*. En cela, les deux traitements de l'incertitude se distinguent ensemble de celui qui est offert par les logiques non-monotones. Toutefois, au-delà de cette première convergence, le traitement de l'incertitude par les réseaux bayésiens diffère de celui qui est à l'oeuvre dans les systèmes tels MYCIN par deux aspects fondamentaux. En premier lieu, le concept numérique sur lequel repose le traitement de l'incertitude dans les réseaux bayésiens est probabiliste. Du coup, il rend disponible toute la théorie classique des probabilités.

En second lieu, mais surtout, ce concept numérique n'est pas manipulé de la même façon dans MYCIN et dans un réseau bayésien. Dans MYCIN, les facteurs de certitude sont attribués à des hypothèses sur la base de règles locales du type de celle que nous venons de mentionner. Le facteur de certitude attaché à une hypothèse découle des informations sous une règle locale. Dans cette mesure, on peut considérer que les facteurs de certitude généralisent la notion de degré de vérité. A l'inverse, dans les réseaux bayésiens, les états du monde ont d'emblée des probabilités, qui varient en fonction des informations obtenues et selon le principe général de la conditionalisation bayésienne. La variation des probabilités est globale au sens où une nouvelle information implique la révision de toutes les probabilités. Pearl parle d'un traitement *sémantique* de l'incertitude dans les réseaux bayésiens, et l'oppose au traitement *syntaxique*[3] qui est à l'oeuvre dans un système expert du même type que MYCIN.[4]

[2] Shortliffe et Buchanan (1984) p. 239.

[3] Pearl (1988) p. 3.

[4] Il convient de souligner ici qu'il n'y a pas d'implication du caractère syntaxique

Le principal attrait des traitements syntaxiques de l'incertitude est computationnel. Du caractère local des règles de la base de connaissance, il découle en effet qu'il est possible de définir une procédure modulaire pour déterminer le facteur de certitude associé au conséquent de la règle. La principale contre-partie de cette facilité computationnelle réside dans la nécessité de définir des règles très nombreuses pour prendre en compte les exceptions aux règles.[5] A l'inverse, les traitements sémantiques se heurtent à l'obstacle computationnel :

> puisque la syntaxe n'indique aucune procédure utile, nous devons construire des mécanismes spéciaux pour convertir les entrées déclaratives en des routines pour répondre à des questions.[6]

Cela suppose de contourner l'obstacle du caractère global des conséquences de l'acquisition d'information dans le contexte sémantique.

Dans les réseaux bayésiens, ce contournement s'effectue au moyen d'un « truc » :

> Le truc, dès lors, est d'encoder les connaissances de telle sorte que ce qu'on peut ignorer est reconnaissable (*the ignorable is recognizable*) ou, mieux encore, que ce qu'on peut ignorer est identifié rapidement et accessible facilement.[7]

L'encodage qui convient est graphique. Ainsi, les réseaux bayésiens comportent des graphes sur lesquels on peut lire ce qui peut être ignoré et, positivement, ce qui doit être pris en compte à l'occasion de la révision de la probabilité d'un état du monde donné. Il apparaît alors que la composante graphique des réseaux bayésiens est essentielle au traitement sémantique de l'incertitude qu'ils véhiculent. A l'inverse, les graphes qui sont mobilisés dans le contexte syntaxique sont toujours des auxiliaires et jamais porteurs d'informations indispensables pour mener l'inférence.

Nous avons décrit le paysage théorique dans lequel les réseaux bayésiens apparaissent. Plus précisément, nous avons qualifié la réponse que les réseaux bayésiens apportent à la question du traitement de l'incertitude en intelligence artificielle. Il nous reste à comprendre les détails de cette réponse. Cela

---

du traitement de l'incertitude dans MYCIN au caractère non probabiliste du concept numérique utilisé pour traiter l'incertitude. Ainsi PROSPECTOR est-il un système expert allié stratégie syntaxique et probabilités. Réciproquement, des concepts numériques non probabilistes peuvent être utilisés dans le cadre d'un traitement sémantique de l'incertitude.

[5] Pour une présentation des autres contre-parties de l'intérêt computationnel des approches syntaxiques, voir Pearl (1988) sous-section 1.2.2.

[6] Pearl (1988) p. 12.

[7] Pearl (1988) pp. 12–13.

ne sera possible qu'après avoir défini rigoureusement les réseaux bayésiens et présenté les résultats fondamentaux qui les concernent. Nous le faisons dans les deux prochaines sous-sections.

1.1.2 Que sont les réseaux bayésiens ? Définitions

Les réseaux bayésiens sont des couples composés d'un graphe orienté acyclique et d'une distribution de probabilités, définis sur un même ensemble de variables, et qui entretiennent un certain rapport. Définir complètement les réseaux bayésiens requiert donc en premier lieu d'introduire des notions de théorie des graphes d'une part et des notions probabilistes d'autre part. Nous le faisons dans l'appendice au présent chapitre. Les définitions que nous donnons dans cet appendice, ainsi que la terminologie que nous y introduisons, sont usuelles. Mais définir les réseaux bayésiens, c'est ensuite et surtout définir le rapport qu'entretiennent les deux éléments du couple qu'un réseau bayésien compose. C'est sur ce point que nous nous concentrons dans le corps du chapitre.

Notations. Dans la suite de la présente sous-section, $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$ est un ensemble fini de variables aléatoires dont chacune admet un nombre de valeurs fini.[8]
$<$ est un ordre strict sur $\mathbf{V}$. On suppose que les indices des variables de $\mathbf{V}$ correspondent à leur ordonnancement pour $<$: $V_1 < V_2 < \ldots < V_n$.
p une distribution de probabilités sur $\mathbf{V}$ et G un graphe orienté acyclique sur $\mathbf{V}$.
La notion à définir est celle de réseau bayésien (G, p) sur $\mathbf{V}$. Mais avant d'en arriver là, et pour que la définition énoncée fasse sens, il convient toutefois de définir la notion – uniquement probabiliste – de parent markovien.

1.1.2.1 Parents markoviens d'une variable

Définition 1.1 (Parents markoviens) *Soit V_i une variable de $\boldsymbol{V}$.*
Un ensemble $\boldsymbol{PM}_i$ de parents markoviens *de V_i pour p et $<$ dans $\boldsymbol{V}$ est sous-ensemble de $\boldsymbol{V}$ minimal parmi ceux qui ont les propriétés suivantes :*
- *tous les éléments de $\boldsymbol{PM}_i$ sont des prédécesseurs de V_i pour $<$;*
- *V_i est indépendant pour p de $\{V_1, V_2, \ldots, V_{i-1}\} \setminus \boldsymbol{PM}_i$ relativement à $\boldsymbol{PM}_i$.*

[8] Tous les ensembles de variables aléatoires que nous considérerons sont finis et tels que chaque variable est discrète et peut prendre un nombre fini de valeurs. Nous ne connaissons pas de travaux dans lesquels l'une ou l'autre de ces restrictions est levée. Nous ne mentionnerons plus ces restrictions dans la suite.

Pour le dire autrement et de manière plus imagée, les variables de $\mathbf{PM}_i$ suffisent exactement à rendre les autres variables de $\{V_1, \ldots, V_{i-1}\}$ non pertinentes pour la probabilité des valeurs de V_i. De façon similaire, dans une chaîne de Markov, un état suffit à rendre non pertinents tous les états qui le précèdent relativement à l'état qui lui succède immédiatement. On comprend alors l'utilisation de l'adjectif « markovien » dans le contexte auquel nous nous référons.

A titre d'illustration, on peut considérer une suite infinie de lancers indépendants d'un dé équilibré et définir quatre variables W, X, Y et Z, dans cet ordre, dont les valeurs varient avec le lancer considéré de la façon suivante : pour le n-ième lancer,
W prend pour valeur le résultat du $n+1$-ième lancer ;
X prend pour valeur la somme des résultats des n-ième et $n+1$-ième lancers ;
Y prend pour valeur le résultat du $n+2$-ième lancer ;
Z prend pour valeur la somme des résultats des n-ième, $n+1$-ième et $n+2$-ième lancers.
Sous ces définitions, la valeur de Z ne dépend que des valeurs que prennent X et Y. Z est donc indépendant de W relativement à $\{X, Y\}$. D'un autre côté, la valeur de Z ne dépend ni de celle de X seul, ni de celle de Y seul ; Z n'est ni indépendant de $\{Y, W\}$ relativement à X, ni indépendant de $\{X, W\}$ relativement à Y. $\{X, Y\}$ est donc un ensemble de parents markoviens de Z pour l'ordre que nous avons adopté et la distribution de probabilités p sur $\{W, X, Y, Z\}$[9]. On remarquera que c'est le seul.

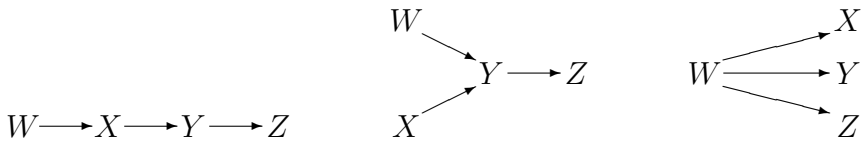
L'idée qui conduit des parents markoviens aux réseaux bayésiens est la suivante : représenter les ensembles de parentalité markovienne par des graphes orientés acycliques. Plus précisément, la représentation consiste à faire des parents markoviens d'une variable, ses parents dans un graphe orienté acyclique. Pour donner forme rigoureuse à cette idée – et donc définir les réseaux bayésiens –, deux concepts doivent encore être introduits : d'abord celui d'accord entre un ordre strict et un graphe orienté acyclique définis sur le même ensemble de variables, et ensuite celui de représentation d'une distribution de probabilités par un graphe orienté acyclique.

1.1.2.2 Accord d'un ordre strict avec un graphe orienté acyclique

Définition 1.2 (Accord de $<$ avec G) $<$ s'accorde *avec G si :*
pour tout V_i et V_j de $\boldsymbol{V}$, si V_i est un ancêtre de V_j, alors $V_i < V_j$.

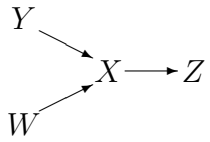
[9] La question de l'interprétation des probabilités n'est pas posée ici. Nous parlons de la distribution de probabilités « naturelle » sur cet ensemble, qu'on peut considérer comme des rapports entre les cas favorables et les cas possibles.

« Ancêtre » a ici le sens graphique défini dans la sous-section 1.4 de l'appendice au présent chapitre : V_i est un ancêtre de V_j dans G si et seulement s'il existe un chemin orienté de V_i à V_j. A titre, maintenant, d'illustration, revenons à l'exemple introduit plus haut. Pour l'ordre sur $\{W, X, Y, Z\}$ que nous avons considéré dans le paragraphe précédent, on trouve les trois graphes orientés acycliques suivants parmi les graphes qui s'accordent avec lui :



G_1 G_2 G_3

Cette liste n'est pas exhaustive. Pour donner une idée de tous les graphes qu'elle comprend, considérons le graphe suivant :



G_4

G_4 ne s'accorde pas avec l'ordre $W < X < Y < Z$ parce que Y est un ancêtre de X dans G_4.

1.1.2.3 Représentation d'une distribution de probabilités par un graphe orienté acyclique

Définition 1.3 (Représentation de p par G) *G* représente *p si : pour toute variable V_i de $\boldsymbol{V}$, l'ensemble $\boldsymbol{PA}(V_i)$ des parents graphiques de V_i constitue un ensemble de parents markoviens de V_i relativement à p et à tout ordre strict sur $\boldsymbol{V}$ qui s'accorde avec G.*

Quand G représente p, on dit aussi que G et p sont *compatibles*, ou encore que p est *markovienne relativement à G*.

Revenons, maintenant, à l'exemple introduit un peu plus haut. Nous avons vu que, pour l'ordre strict que nous avons considéré et la distribution de probabilités sur $\{W, X, Y, Z\}$, un ensemble de parents markoviens de Z est $\{X, Y\}$. Par ailleurs il nous semble clair qu'un ensemble de parents markoviens pour W est $\emptyset$, qu'un ensemble de parents markoviens pour X est

$\{W\}$ et qu'un ensemble de parents markoviens pour Y est $\emptyset$[10]. Dès lors, la distribution de probabilités p sur $\{W, X, Y, Z\}$ est représenté par le graphe orienté acyclique 2. de l'énumération proposée dans le dernier paragraphe.

1.1.2.4 Réseau bayésien

Définition 1.4 (Réseau bayésien) *Un* réseau bayésien *sur un ensemble de variables* $\boldsymbol{V}$ *est un couple* (G, p) *tel que :*

1. *G est un graphe orienté acyclique sur $\boldsymbol{V}$;*
2. *p est une distribution de probabilités sur $\boldsymbol{V}$;*
3. *G représente p.*

D'après ce qui précède, le couple composé du graphe G_2 ci-dessus et de la distribution de probabilités sur $\{W, X, Y, Z\}$ est un réseau bayésien.

1.1.3 Deux résultats fondamentaux

Les deux résultats que nous présentons dans cette sous-section sont fondamentaux au moins au sens où ils portent sur la notion de représentation d'une distribution de probabilités par un graphe qu'on trouve au fondement de la définition 1.4 des réseaux bayésiens.

1.1.3.1 Condition de Markov

Le premier des résultats que nous présentons consiste dans une équivalence entre la représentation d'une distribution de probabilités par un graphe et une autre condition, appelée « condition de Markov ». Ce résultat est le suivant :

Proposition 1.1 (Pearl, 1988) *Soit* $\boldsymbol{V}$ *un ensemble de variables, G un graphe orienté acyclique sur* $\boldsymbol{V}$ *et p une distribution de probabilités sur* $\boldsymbol{V}$*. G représente p si et seulement si chaque variable de* $\boldsymbol{V}$ *est indépendante pour p de tous ses non-descendants dans G relativement à ses parents dans G.*[11]

Ainsi que nous l'avons déjà partiellement indiqué, on appelle « condition de Markov parentale » ou plus simplement « condition de Markov » la condition nécessaire et suffisante de représentation d'une distribution de probabilités par un graphe qui est énoncée dans la proposition 1.1 :

[10] Dans tous les cas, l'ensemble de parents markoviens que nous donnons est en fait unique.

[11] On convient qu'une variable n'appartient pas à l'ensemble de ses non-descendants graphiques.

Définition 1.5 (Condition de Markov) *Soit* $\boldsymbol{V}$ *un ensemble de variables,* G *un graphe orienté acyclique sur* $\boldsymbol{V}$ *et* p *une distribution de probabilités sur* $\boldsymbol{V}$*.*
Le couple (G, p) *satisfait la condition de Markov si toute variable de* $\boldsymbol{V}$ *est indépendante de tous ses non-descendants dans* G *relativement à l'ensemble de ses parents dans* G*.*

Étant équivalente à la condition de représentation mobilisée dans la définition 1.4, elle peut lui être substituée pour produire une définition alternative des réseaux bayésiens.[12] La condition de Markov fait l'objet d'une grande partie des discussions contemporaines portant sur les réseaux bayésiens en général et les réseaux bayésiens causaux en particulier.

1.1.3.2 *d*-séparation

Le second des résultats que nous présentons dans cette sous-section explore la correspondance entre un graphe orienté acyclique et les distributions de probabilités qu'il représente.

_d_-séparation d'un chemin.

Définition 1.6 (*d*-séparation d'un chemin) *Dans un graphe orienté acyclique* G *sur un ensemble de variables* $\boldsymbol{V}$*, un chemin* c *est* d-séparé *par un sous-ensemble* $\boldsymbol{W}$ *de* $\boldsymbol{V}$ *si l'une des deux propositions suivantes est vraie :*

1. *c contient une chaîne $V_i \longrightarrow V_j \longrightarrow V_k$ ou une fourche $V_i \longleftarrow V_j \longrightarrow V_k$ telle que V_j appartient $\boldsymbol{W}$;*
2. *c contient une fourche inversée $V_i \longrightarrow V_j \longleftarrow V_k$ telle que ni V_j, ni aucun de ses descendants n'appartient à $\boldsymbol{W}$.*

A titre d'illustration, notons les d-séparations suivantes dans les graphes G_1, G_2 et G_3 ci-dessus :
Dans G_1, le chemin qui va de W à Z est d-séparé par $\{Y\}$ en vertu du premier disjoint de la clause 1.
Dans G_3, le chemin entre X et Z est d-séparé par $\{W, Y\}$ en vertu du second disjoint de la clause 1.
Dans G_2, le chemin entre W et Y est d-séparé par $\{X\}$ en vertu de la clause 2.

[12] La condition de Markov est utilisée pour définir les réseaux bayésiens par plusieurs auteurs. Voir par exemple Williamson (2005) pp. 14-16.

***d*-séparation de deux ensembles de variables.** La définition 1.6 donnée dans le paragraphe précédent nous permet de définir une relation ternaire de d-séparation entre ensembles de variables d'un graphe orienté acyclique :

Définition 1.7 *Soit un graphe orienté acyclique G sur un ensemble de variables $\boldsymbol{V}$ et $\boldsymbol{W}$, $\boldsymbol{X}$ et $\boldsymbol{Y}$ trois sous-ensembles de $\boldsymbol{V}$.*
$\boldsymbol{Y}$ d-sépare *$\boldsymbol{W}$ et $\boldsymbol{X}$ dans G si tout chemin d'une variable de $\boldsymbol{W}$ à une variable de $\boldsymbol{X}$ est* d-*séparé par $\boldsymbol{Y}$.*

Revenons, à nouveau, à notre exemple. Dans le graphe G_2, $\{W\}$ et $\{Y\}$ sont d-séparés par $\{X\}$ en vertu de ce qui a été mis en évidence dans le paragraphe pécédent. Dans ce même graphe, on notera aussi – et entre autres – la d-séparation de $\{W, Z\}$ par $\{X, Y\}$.

On notera que la notion de d-séparation, de même que celle d'indépendance probabiliste relative, est symétrique : $\mathbf{Y}$ d-sépare $\mathbf{W}$ de $\mathbf{X}$ dans G si et seulement si $\mathbf{Y}$ d-sépare $\mathbf{X}$ de $\mathbf{W}$ dans G.

Second résultat fondamental relatif aux réseaux bayésiens. La notion de d-séparation entre ensembles de variables permet d'énoncer une propriété importante de la correspondance entre un graphe orienté acyclique et les distributions de probabilités qu'il représente :

Théorème 1.1 (Verma et Pearl, 1988) *Soit G un graphe acyclique orienté sur un ensemble de variables $\boldsymbol{V}$ et soit $\boldsymbol{W}$, $\boldsymbol{X}$ et $\boldsymbol{Y}$ trois sous-ensembles de $\boldsymbol{V}$.*
$\boldsymbol{W}$ et $\boldsymbol{X}$ sont d-*séparés par $\boldsymbol{Y}$ dans G si et seulement si $\boldsymbol{W}$ est indépendant de $\boldsymbol{X}$ relativement à $\boldsymbol{Y}$ pour toute distribution de probabilités représentée par G.*

La d-séparation dans un graphe G correspond donc exactement à l'indépendance probabiliste pour toutes les distributions de probabilités représentées par G. Il en découle immédiatement que la d-séparation dans le graphe G d'un réseau bayésien (G, p) implique l'indépendance probabiliste pour p. Ainsi, la d-séparation dans le graphe G_2 ci-dessus implique l'indépendance probabiliste relative pour la distribution de probabilités p impliquée par la description proposée pour la situation. Parce que (G_2, p) est un réseau bayésien, la propriété graphique de d-séparation dans G_2 devient un critère d'indépendance probabiliste relative pour p.

Dans les deux sous-sections qui s'achèvent ici, nous avons défini les réseaux bayésiens et présenté deux résultats fondamentaux les concernant.

Armés de cela, nous pouvons revenir à la question du traitement de l'incertitude. Plus généralement, nous pouvons maintenant en venir aux applications qu'autorise la notion de réseaux bayésiens. En d'autres termes, nous en venons maintenant aux utilisations des réseaux bayésiens.

1.1.4 Utilisations des réseaux bayésiens

Ainsi que nous l'avons annoncé plus haut (dans la sous-section 1.1.1), les réseaux bayésiens permettent de réviser les probabilités attribuées à des états du monde dans un contexte d'incertitude. Cela, toutefois, n'est possible que parce qu'ils constituent des représentations particulièrement économiques de distributions de probabilités. Dans ces conditions, nous élucidons ce dernier point avant d'expliquer comment s'actualisent les probabilités dans un réseau bayésien.

1.1.4.1 Définitions économiques de distributions de probabilités

Les réseaux bayésiens comme définitions de distributions de probabilités. Un réseau bayésien non interprété permet en premier lieu de définir la distribution de probabilités qui le compose. Pour le comprendre, il nous faut présenter d'abord un résultat élémentaire du calcul des probabilités :

Proposition 1.2 (Règle de la chaîne) *Pour tout ensemble de variables aléatoires* $\boldsymbol{V} = (V_1, V_2, \ldots, V_n)$*, toute distribution de probabilités* p *sur* $\boldsymbol{V}$ *et toute valeur* $(v_1, v_2, \ldots, v_n)$ *de* $(V_1, V_2, \ldots, V_n)$*,*
$p(v_1, v_2, \ldots, v_n) = p(v_1).\prod_{i=2}^{n} p(v_i|v_1, \ldots, v_{i-1})$*.*

Considérons maintenant un ordre strict $<$ et une distribution de probabilités p sur $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$. Pour tout $1 < i < n$, on note $\mathbf{pm}_i$ un ensemble de parents markoviens de V_i pour p et $<$. D'après la définition 1.1 des parents markoviens d'une variable, l'égalité énoncée par la règle de la chaîne se simplifie alors en :

$$p(v_1, v_2, \ldots, v_n) = \prod_{i=1}^{n} p(v_i|\mathbf{pm}_i). \tag{1.1}$$

De l'équation (1.1) il découle qu'une distribution de probabilités p sur un ensemble de variables $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$ est complètement définie par :

1. un ordre strict $<$ sur $\mathbf{V}$;
2. pour chaque variable V_i de $\mathbf{V}$, un ensemble de parents markoviens $\mathbf{PM}_i$ pour p et $<$;

3. les probabilités conditionnelles $p(v_i|\mathbf{pm}_i)$ pour $1 \leq i \leq n$, $v_i \in Val(V_i)$ et $\mathbf{pm}_i \in Val(\mathbf{PM}_i)$.

Rappelons, maintenant, qu'un graphe G qui représente p donne, pour chaque variable V_i de $\mathbf{V}$, un ensemble de parents markoviens de V_i pour p et tout ordre strict compatible avec G. Plus précisément, les parents graphiques dans G d'une variable de $\mathbf{V}$ sont un ensemble de parents markoviens de cette variable pour p et tout ordre strict sur V compatible avec G. Il en découle qu'une distribution de probabilités p sur $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$ est complètement définie par :

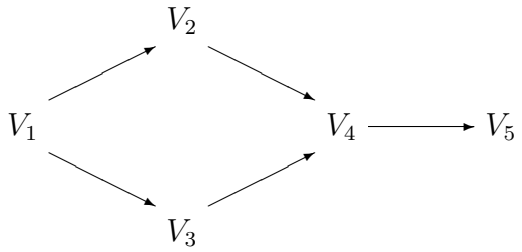
1. un graphe orienté acyclique G qui représente p ;
2. l'ensemble de probabilités conditionnelles $\mathbf{PC} = \{p(v_i|\mathbf{pa}_i)$ pour $1 \leq i \leq n$, $v_i \in Val(V_i)$ et $\mathbf{pa}_i \in Val(\mathbf{PA}_i)\}$.[13]

Dans ces conditions, d'une part le réseau bayésien (G, p) est complètement défini par G et $\mathbf{PC}$, et d'autre part (G, p) ainsi défini constitue lui-même une définition de p – au sens où il permet de spécifier $p(\mathbf{v})$ pour toute valeur $\mathbf{v}$ de $\mathbf{V}$.

Caractère économique de la définition. La définition d'une distribution de probabilités p par un réseau bayésien (G, p) selon les modalités que nous venons d'indiquer est intéressante dans la mesure où elle est particulièrement économique. En effet, si l'ensemble $\mathbf{V}$ sur lequel est défini (G, p) est $\{V_1, V_2, \ldots, V_n\}$, alors l'ensemble de probabilités conditionnelles $\mathbf{PC}$ compte $\sum_{i=1}^{n}(\|Val(V_i)\|.\|Val(\mathbf{PA}(V_i))\|)$ éléments. En tenant compte de la contrainte exprimée par la proposition 1.6 énoncée dans l'appendice, il faut, pour spécifier complètement $\mathbf{PC}$, $\sum_{i=1}^{n}[(\|Val(V_i)\|-1).\|Val(\mathbf{PA}(V_i))\|]$ paramètres. En d'autres termes, $\sum_{i=1}^{n}[(\|Val(V_i)\| - 1).\|Val(\mathbf{PA}(V_i))\|]$ paramètres sont nécessaires pour définir p une fois qu'on a G. De l'autre côté, sans (G, p) mais toujours en tenant compte de la contrainte exprimée par la proposition 1.6, il faut $\prod_{i=1}^{n}\|Val(V_i)\| - 1$ paramètres pour définir la même distribution p. Or $\prod_{i=1}^{n}\|Val(V_i)\| - 1$ est strictement supérieur à $\sum_{i=1}^{n}[(\|Val(V_i)\|-1).\|\mathbf{PA}(V_i)\|]$ dès qu'il existe un couple de variables (V_i, V_j) qui ne sont pas adjacentes dans le graphe représentant p qu'on considère.

A titre d'illustration, que nous reprenons à Williamson[14], une distribution de probabilités p sur un ensemble $\{V_1, V_2, V_3, V_4, V_5\}$ de variables aléatoires binaires est définie par $2^5 - 1 = 31$ paramètres, mais le fait de connaître le graphe G :

[13] Rappelons que nous notons $\mathbf{PA}_i$ l'ensemble des parents graphiques de la variable V_i.
[14] Williamson (2005) p.19.



qui représente p permet de définir p au moyen de $1 + 2 + 2 + 4 + 2 = 11$ paramètres seulement. Williamson montre que dans le cas général une distribution de probabilités p représentée par un graphe orienté acyclique sur n variables dont chacune a au plus k parents et K valeurs peut être définie par au plus $n.K^k.(K-1)$ paramètres. La fonction qui à n associe le nombre de paramètres nécessaires à la spécification de p est alors linéaire. Sans graphe représentant p, cette même distribution de probabilités est définie par un nombre de paramètres de l'ordre de K^n ; la fonction qui à n associe le nombre de paramètres nécessaires à la spécification de p est alors exponentielle.

Un réseau bayésien (G, p) défini par G et par l'ensemble de probabilités conditionnelles **PC** constitue donc une définition remarquablement économique de p. Selon la proposition 1.1, toute distribution de probabilités est représentée par au moins un graphe orienté acyclique. Il en découle que toute distribution de probabilités peut être définie selon les voies économiques que nous venons de décrire. Dans tous ces cas, l'intérêt qu'il y a intrinsèquement à disposer d'une définition économique pour une distribution de probabilités donnée se double de l'intérêt extrinsèque correspondant à la possibilité de fonder sur cette définition des méthodes efficaces d'actualisation des probabilités.

1.1.4.2 Actualisation des probabilités dans un réseau bayésien

En vue de comprendre montrer que les réseaux bayésiens constituent des outils pour l'actualisation des probabilités, nous commençons ici par prendre un exemple. Soit $\{A, B\}$ un ensemble de deux variables binaires et supposons que la distribution des probabilités p sur $\{A, B\}$ est définie par :

$$p(a_1, b_1) = \tfrac{1}{6} \qquad p(a_1, b_2) = \tfrac{1}{12}$$

$$p(a_2, b_1) = \tfrac{3}{16} \qquad p(a_2, b_2) = \tfrac{9}{16}$$

Supposons encore qu'on apprend que B prend la valeur b_2[15]. Une question qui se pose alors est la suivante : quelle probabilité doit-on accorder à l'événement

[15]Du point de vue de l'interprétation de cette situation, deux scénarios sont envisageables : soit p mesure des degrés de croyance subjectifs et apprendre que B prend la

que A prend la valeur a_1 (resp. prend la valeur a_2) ? De façon plus générale, le problème est le suivant : comment se modifie une distribution de probabilités p sur un ensemble de variables $\mathbf{V}$ à la lumière de l'information I selon laquelle un sous-ensemble $\mathbf{W}$ de $\mathbf{V}$ prend la valeur $\mathbf{w}$?

Nous connaissons une réponse au problème que nous venons de présenter : la probabilité p' est la probabilité conditionnelle $p(\_|w)$, où p est la distribution de probabilités physique initialement connue et w l'information obtenue à prendre en compte. Dans l'exemple que nous avons proposé, on obtient ainsi pour a_1 :

$$p'(a_1) = p(a_1|b_2) = \frac{p(a_1, b_2)}{p(b_2)} = \frac{p(a_1, b_2)}{p(a_1, b_2) + p(a_2, b_2)} = \frac{\frac{1}{12}}{\frac{1}{12} + \frac{9}{16}} = \frac{4}{31}.$$

On calculerait de façon similaire les probabilités des autres valeurs des sous-ensembles de $\{A, B\}$.

Ce mode de calcul, toutefois, est peu pratiquable. Plus précisément : quand les situations envisagées se complexifient, il devient rapidement impossible d'actualiser les probabilités selon les voies que nous avons empruntées pour calculer $p'(a_1)$. Si, en revanche, on connaît un graphe qui représente la distribution de probabilités initiale, les choses deviennent plus faciles – et, surtout, effectivement traitables. En effet, étant donné un graphe G qui représente une distribution de probabilités sur $\mathbf{V}$, la valeur d'une variable A de $\mathbf{V}$ varie avec les seules variations des valeurs de ses parents directs dans G. Il en découle que disposer d'un graphe qui représente économiquement p (et donc p') autorise à simplifier les calculs d'actualisation de probabilités selon les voies indiquées par l'équation 1.1. En outre, cette représentation autorise une simplification des opérations algébriques d'actualisation des probabilités.[16] Au total, les réseaux bayésiens constituent non seulement des définitions économiques de distributions de probabilités, mais encore des outils d'actualisation des distributions de probabilités dont ils sont des définitions. L'actualisation des probabilités qu'ils autorisent repose toujours sur la conditionalisation bayésienne ; nous y voyons une des raisons pour lesquelles on parle de réseaux *bayésiens*.

Ainsi que nous l'avons indiqué plus haut, l'apparition des réseaux bayésiens est inséparable de leur utilisation à des fins d'actualisation des probabilités. Plus précisément maintenant, Pearl a développé très tôt[17] un

valeur b_2 est une information nouvelle sur le monde, soit p est une distribution de probabilités génériques objective pour une classe d'individus et apprendre que B prend la valeur b_2 est une information qui résulte de la focalisation sur un des individus de cette classe.

[16] Sur ce point voir Lauritzen et Spiegelhalter (1988) p. 165.

[17] Dans Pearl (1982).

algorithme d'actualisation des probabilités dans les graphes bayésiens qui sont des arbres – c'est-à-dire qui sont tels que pour toute variable du graphe sauf une (appelée racine), il existe exactement une flèche qui pointe vers elle. A partir de ce travail fondateur, le problème de l'actualisation automatique des probabilités a été résolu pour des classes de graphes de plus en plus vastes. Une solution pour le cas général est présentée dans Lauritzen et Spiegelhalter (1988).

Les utilisations des réseaux bayésiens que nous venons de présenter reposent toutes deux sur ceci que connaître un graphe orienté acyclique représentant une distribution de probabilités réduit considérablement le nombre de paramètres nécessaires à la définition de cette distribution. Ainsi que nous l'avons expliqué déjà dans la sous-section 1.1.1, la composante graphique des réseaux bayésiens est essentielle à ce point. En vue de compléter notre réponse à la question de savoir à quoi servent les réseaux bayésiens, il convient donc de discuter plus précisément le statut des graphes qui composent les réseaux bayésiens.

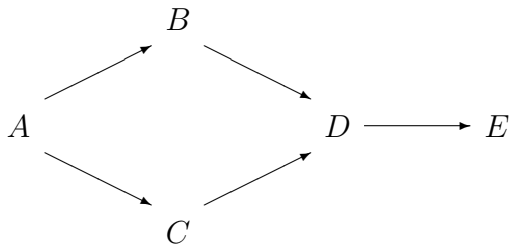
1.1.5 Graphes bayésiens

1.1.5.1 Statut des graphes bayésiens

Dans un réseau bayésien (G, p) sur un ensemble de variables $\mathbf{V}$, G et p sont compatibles. On peut donc considérer que G indique un ensemble de parents markoviens de chacune des variables de $\mathbf{V}$ pour p et pour tout ordre strict sur $\mathbf{V}$ qui s'accorde avec G – et c'est d'ailleurs sur cette lecture qu'est fondée l'analyse du paragraphe précédent. Cette information peut être aisément exprimée sous la forme d'une affirmation portant sur le seul p. Nous le montrons dans un cas particulier.

Considérons un ensemble $\{A, B, C, D, E\}$ de cinq variables binaires de valeurs respectives a_1 et a_2, b_1 et b_2, ..., e_1 et e_2, et un réseau bayésien sur cet ensemble défini par :

1. le graphe G suivant :



2. la spécification des probabilités conditionnelles suivantes (dont on a vu qu'elles suffisent à déterminer complètement p en présence de G) :

$p(a_i)$	pour $1 \leq i \leq 2$
$p(b_i \mid a_j)$	pour $1 \leq i, j \leq 2$
$p(c_i \mid a_j)$	pour $1 \leq i, j \leq 2$
$p(d_i \mid b_j, c_k)$	pour $1 \leq i, j, k \leq 2$
$p(e_i \mid d_j)$	pour $1 \leq i, j \leq 2$

L'information dont G est porteur dans ce contexte peut être exprimée par l'expression linguistique suivante : « pour tout ordre strict $<$ sur $\{A, B, C, D, E\}$ tel que $A < B$, $A < C$, $B < D$, $C < D$ et $D < E$, une suite de parents markoviens pour p et $<$ des variables (A, B, C, D, E) est $(\emptyset, \{A\}, \{A\}, \{B, C\}, \{D\})$ ». Plus généralement, il est donc bien toujours possible d'exprimer dans le langage naturel l'information dont un graphe bayésien est porteur. Cette information n'exige ni une représentation réticulaire, ni même une représentation graphique. L'exemple par lequel nous illustrons ce fait montre en outre que la représentation de cette information par une expression du langage naturel peut toujours être menée à bien en suivant une procédure mécanique très simple ; sa représentation graphique est donc toujours dispensable en fait. Or, c'est en tant que les graphes bayésiens sont porteurs de cette information que les réseaux bayésiens peuvent servir à définir de façon économique les distributions de probabilités. La question de ce à quoi les réseaux bayésiens servent se pose donc à nouveaux frais, en même temps qu'elle se précise. Il s'agit en effet maintenant de savoir à quoi servent les *graphes* bayésiens.

1.1.5.2 A quoi servent les graphes bayésiens ?

Nous envisageons deux réponses à la question que nous posons ici. Selon la première, l'information dont un graphe bayésien est porteur est plus facilement appréhendée si elle est représentée sous cette forme graphique plutôt que par une expression dans le langage naturel. Cette réponse nous semble à la fois facile à accepter et difficile à justifier ; nous ne nous y attardons pas. Selon la seconde réponse que nous envisageons, la représentation graphique de l'information exprimée par le graphe G d'un réseau bayésien (G, p) sur $\mathbf{V}$ rend visibles des propriétés non triviales de p. Le théorème 1.1 indique en effet que la d-séparation dans G constitue un critère nécessaire et suffisant d'indépendance probabiliste pour toute distribution de probabilités compatible avec G – et donc en particulier pour p. Sur G, on lit toutes les indépendances probabilistes pour p – relatives ou non – qui découlent de l'information relative à p que G représente.

Dans ces conditions, on peut considérer un graphe bayésien comme une

construction qui intervient dans la résolution du problème suivant : étant donnés un ensemble de variables aléatoires **V**, un ordre strict $<$ et une distribution de probabilités p sur **V**, quelles sont les indépendances probabilistes qui découlent de la donnée d'une suite **PM** d'ensembles de parents markoviens des variables de **V** pour p et $<$?

Bien sûr, on pourrait identifier toutes les indépendances de ce type sans recourir à un graphe G représentant p. On raisonnerait sur les indépendances probabilistes en prenant pour prémisses les indépendances probabilistes dont **PM** nous informe explicitement et pour règles d'inférence les propriétés de la relation ternaire d'indépendance probabiliste relative – dont les principales sont bien connues.[18] Pour tout triplet (**W**, **X**, **Y**) de sous-ensembles de **V**, une démonstration de ce type permet d'établir, si c'est le cas, que les indépendances probabilistes dont **PM** nous informe explicitement impliquent que **W** est indépendant de **X** relativement à **Y**. Toutefois, une telle démonstration ne permet pas d'établir qu'une indépendance probabiliste ne découle pas des indépendances probabilistes explicitement véhiculées par la donnée de **PM**. En outre, même dans le cas favorable où il y a effectivement indépendance probabiliste, une démonstration à partir des indépendances véhiculées par **PM** et des propriétés de l'indépendance probabiliste consiste en des inférences qui peuvent être nombreuses et qu'il peut être difficile d'identifier et fastidieux de mettre en oeuvre.

A l'inverse, le recours à un graphe G représentant p permet, pour tout triplet (**W**, **X**, **Y**), de déterminer si les indépendances probabilistes données impliquent ou non que **W** et **X** sont indépendants relativement à **Y**. La démonstration proprement dite de la réponse à cette question est réduite à la portion congrue une fois G connu : elle consiste en effet tout entière dans la lecture de la d-séparation ou non de X et Y par Z dans G. Ainsi, pour reprendre l'exemple de la suite de lancers d'un dé, on lit sur G_2 les indépendances et les dépendances probabilistes relatives pour p. A titre d'illustration, il apparaît sur G_2 que, pour p, W est indépendant de Z relativement à $\{X, Y\}$. Il apparaît à l'inverse qu'il existe une distribution de probabilités compatibles (qui n'est pas forcément p) avec G_2 pour laquelle X et Z ne sont pas indépendants relativement à Y.

Il est apparu que les réseaux bayésiens, considérés dans cette première section comme des objets formels, permettent de définir de façon économique une distribution de probabilités sur un ensemble fini de variables aléatoires susceptibles de prendre chacune un nombre fini de valeurs. Cette propriété ne dépend que de la représentation par tout graphe bayésien G d'un réseau (G, p)

[18] Sur ce point, voir par exemple Williamson (2005) pp. 16, Pearl (2000) p. 11, Spohn (1994).

d'une certaine information relative à p. Cette information peut toujours être représentée sans recourir à G, mais sa représentation par G non seulement la rend plus facile à appréhender, mais encore rend apparentes les réponses à une classe de questions relatives aux indépendances probabilistes pour p.

Ces propriétés, toutefois, sont rarement mobilisées pour elles-mêmes. En effet, le recours aux réseaux bayésiens se fait presque toujours en référence à une interprétation – c'est-à-dire, plus précisément, à une interprétation des flèches qui figurent dans les graphes qui composent les réseaux bayésiens. Les interprétations envisagables sont multiples. En tant qu'elle s'attache aux propriétés de l'objet formel, la présentation des réseaux bayésiens que nous avons proposée dans cette première section permet de rendre compte de ces interprétations multiples. Cela n'implique évidemment pas que nous nous apprêtons à les discuter chacune à son tour. Positivement, nous concentrons notre attention sur l'interprétation causale des réseaux bayésiens. La deuxième section de ce chapitre est consacrée à expliquer le rapport qui existe entre les réseaux bayésiens et la causalité.

1.2 Réseaux bayésiens et causalité

L'idée qu'on trouve au fondement de l'interprétation causale des réseaux bayésiens est la suivante : considérer que les flèches du graphe orienté qui compose un réseau bayésien représentent les relations de cause à effet directes entre les variables de l'ensemble sur lequel le réseau est défini. Pour le dire plus concisément, un réseau bayésien causal est un réseau bayésien dont le graphe est causal.

En préambule de cette section consacrée aux réseaux bayésiens causaux, il convient de s'arrêter à l'idée selon laquelle les relations de cause à effet que représentent les flèches du graphe d'un réseau bayésien causal sont des relations de cause à effet *directes*. Une relation de cause à effet entre C et E est directe quand l'influence causale de C sur E ne se réduit pas à l'influence sur E d'une variable A elle-même influencée par C. La propriété d'être directe, pour une relation de cause à effet entre variables, est relative à l'ensemble de variables qu'on considère. Soient en effet les variables correspondant pour l'une à la propriété d'être porteur d'un gène de susceptibilité pour le cancer du sein et pour l'autre de recevoir une chimiothérapie. On note ces variables respectivement G et C. Relativement à un ensemble de variables $\mathbf{V}$ auquel appartient la variable A correspondant à la propriété d'être atteint d'un cancer, G ne cause pas directement C. En effet, l'influence de G sur C se réduit à l'influence de A sur C. Mais si ni A, ni aucune variable causalement interposée entre G et C n'appartient à $\mathbf{V}$, alors G cause directement C dans

V. Le fait que le caractère direct d'une relation de cause à effet dépend de l'ensemble de variables considéré est pris en compte dans la définition d'un réseau bayésien causal comme réseau bayésien dont le graphe représente les relations de cause à effet directes *entre les variables de l'ensemble sur lequel le réseau est défini*.

La notion de réseau bayésien causal ainsi définie a pour corrélats :

1. *une hypothèse relative à la causalité*. En effet, parler de réseau bayésien causal n'est possible que si les relations de cause à effet directes au sein d'un ensemble de variables donné peuvent bien être représentées par les flèches qui figurent dans un graphe bayésien sur cet ensemble. Comme cette hypothèse semble porter moins sur la causalité elle-même que sur la façon dont elle peut être représentée, nous parlerons de l'*hypothèse de représentation* ;
2. *une hypothèse relative au rapport entre la causalité et les probabilités*, selon laquelle le graphe $GC_{\mathbf{V}}$ qui représente les relations de cause à effet directes entre les variables d'un ensemble $\mathbf{V}$ représente la distribution de probabilités p sur cet ensemble. En d'autres termes, le couple $(GC_{\mathbf{V}}, p)$ satisfait la condition de Markov (définition 1.5). L'hypothèse est donc que toute variable de $\mathbf{V}$ est indépendante, relativement à l'ensemble de ses causes directes dans $\mathbf{V}$, de toutes les variables de $\mathbf{V}$ dont elle n'est pas un ancêtre causal. Cette hypothèse est connue comme : *condition de Markov causale*.

Conformément à l'objectif annoncé au début du chapitre, l'effort d'analyse que nous faisons dans la présente section se concentre sur ces hypothèses. Pour chacune, nous l'analysons et nous attachons à montrer qu'elle n'est pas problématique. Avant, toutefois, d'en venir là, il convient ici de motiver l'introduction des réseaux bayésiens causaux, c'est-à-dire de présenter leurs principales utilisations. C'est ce que nous faisons dans une première sous-section.

1.2.1 Utilisations des réseaux bayésiens causaux

Pour définir et distinguer deux types d'utilisations des réseaux bayésiens causaux, il convient de s'arrêter à un problème soulevé par la notion non interprétée de réseau bayésien. Ce problème est celui de la construction d'un graphe G représentant une distribution de probabilités donnée, définie sur un ensemble de variables. Ce problème n'est pas trivial ; il est même généralement NP-complet.[19] Toutefois on sait :

[19] On montre par exemple que le problème consistant à construire un réseau bayésien dans lequel chaque variable a au plus K parents et dont la probabilité étant donnée

1. le résoudre de manière approchée pour un espace de recherche restreint. Plus précisément : étant données une mesure de la distance entre distributions de probabilités définies sur le même ensemble de variables et une classe de graphes orientés acycliques, on sait construire le graphe qui, parmi ceux de cette classe, représente la distribution de probabilités la plus proche d'une distribution p donnée [20] ;
2. le résoudre de façon incomplète moyennant l'hypothèse de fidélité :

   Définition 1.8 (Fidélité) *Soit (G, p) un réseau bayésien. Il n'existe pas d'indépendances pour p autres que celles qui sont impliquées par ceci que (G, p) est un réseau bayésien.*[21]

   Sous cette hypothèse, on sait construire une représentation graphique de l'ensemble des graphes orientés acycliques qui représentent une distribution de probabilités donnée. Ce graphe acyclique, partiellement orienté en général, est usuellement appelé « patron » (*pattern*). Les principaux algorithmes qui réalisent la tâche de construction de patrons causaux se sont développés en deux séries parallèles : d'un côté, les algorithmes IC et IC* de Verma et Pearl[22], de l'autre les algorithmes SGS, PC et PC* et CI et FCI de Spirtes, Glymour et Scheines[23].

En rapport avec le problème de la construction d'un graphe orienté acyclique représentant une distribution de probabilités donnée, on peut identifier et distinguer deux types d'utilisations des réseaux bayésiens causaux. Dans les deux cas, les utilisations sont fondées sur ceci, que nous mettrons en évidence plus bas, que les hypothèses corrélatives de la notion de réseau bayésien causal semblent pouvoir être acceptées. Dans le premier cas, on en tire l'idée selon laquelle une méthode rivale de celles que nous venons de présenter consiste à considérer le graphe qui donne une représentation naturelle de la causalité sur un ensemble de variables comme représentant les distributions de probabilités physiques sur cet ensemble. Dans le second cas, on en tire l'idée selon laquelle le graphe qu'on obtient en utilisant les méthodes du paragraphe précédent est une représentation naturelle des relations de cause à effet directes sur l'ensemble de variables considéré. Pour les méthodes de type 2., il s'agit plus précisément d'une représentation naturelle

les données statistiques est supérieure à un réel fixé est un problème NP-complet. Voir Chickering (1996).

[20] Pour une présentation synthétique de la famille des méthodes que nous décrivons, voir Leray (2005). Pour une présentation détaillée d'une méthode de cette famille, voir Williamson (2005) sections 3.5 à 3.11.

[21] Pearl parle de « stabilité » plutôt que de fidélité (Pearl, 2000, p. 48).

[22] Pour une présentation, voir Pearl (2000) sections 2.5 et 2.6.

[23] Pour une présentation, voir Spirtes et al. (1993) sections 5.4 et 6.7.

de l'ensemble des relations de cause à effet directes qu'on peut inférer des relations d'indépendance probabiliste entre les variables de **V**.

Le premier type d'utilisations des réseaux bayésiens causaux est indissociable du contexte d'apparition des réseaux bayésiens que nous avons présenté dans la sous-section 1.1.1. Dans ce premier cas, la causalité est liée aux réseaux bayésiens selon la modalité suivante : connaître les relations de cause à effet directes sur un ensemble de variables facilite la tâche qui consiste à construire un graphe orienté acyclique qui représente une distribution de probabilités physiques sur cet ensemble. En d'autres termes, la causalité directe est considérée comme connue et joue le rôle de guide pour la construction de graphes bayésiens. On trouve une présentation et une discussion de ce type d'utilisations des réseaux bayésiens causaux dans la section 6 de Gillies (2002). Nous aurons à y revenir plus loin dans le chapitre.

Le second type de contextes dans lesquels les réseaux bayésiens causaux apparaissent dérive des méthodes 2. pour la construction de graphes bayésiens. Contrairement aux méthodes 1., les méthodes 2. ne contraignent pas *a priori* la forme du graphe bayésien. Elles sont donc compatibles avec l'idée selon laquelle ce graphe représente la causalité directe sur l'ensemble de variables considéré. Autrement dit, les méthodes de type 2. sont utilisées comme des méthodes de construction de graphes *causaux*. Contrairement aux utilisations du premier type, celles du second type ne supposent donc pas connues les relations de cause à effet directes entre les variables de l'ensemble considéré. Ces relations sont même ce qu'on escompte apprendre – au moins partiellement – de la construction de l'ensemble des graphes orientés acycliques représentant une distribution de probabilités donnée sur un ensemble de variables **V**. En d'autres termes, les réseaux bayésiens causaux apparaissent en second lieu dans un contexte *d'inférence causale*. Le lecteur aura compris que ce second contexte est celui qui nous intéresse dans cette première partie de notre travail. Avant d'en venir spécifiquement à lui, il convient toutefois de proposer une discussion générale de l'hypothèse de représentation d'abord et de la condition de Markov causale ensuite.

1.2.2 L'hypothèse de représentation

L'hypothèse selon laquelle les relations de cause à effet directes au sein d'un ensemble de variables donné peuvent être représentées par les flèches qui figurent dans un graphe bayésien sur cet ensemble se décompose en trois sous-hypothèses :

1. Les *relata* de la causalité peuvent être représentés par des variables.
2. La relation de causalité directe au sein d'un ensemble de variables peut

être représentée par une relation binaire.

3. La relation de causalité directe au sein d'un ensemble de variables peut être représentée par une relation asymétrique.

Nous présentons et discutons chacune à son tour ces trois composantes de l'hypothèse de représentation.

1.2.2.1 Première composante de l'hypothèse de représentation : Représenter les *relata* causaux par des variables

Avant et en vue de discuter cette première composante de l'hypothèse de représentation, il convient de s'arrêter sur un point de terminologie. Telle que nous l'avons utilisée jusqu'ici, l'expression « relation de cause à effet » désigne le couple constitué d'une cause et de son effet. Par « relation de causalité » ou plus simplement « causalité », nous désignons au contraire ce qui est une relation au sens logique du terme. Plus précisément il s'agit de l'ensemble des couples constitués chacun d'une cause et de son effet – c'est-à-dire l'ensemble des relations de cause à effet. Nous pouvons maintenant en venir à la sous-hypothèse 1 elle-même. Cette sous-hypothèse 1 demande à être soigneusement discutée. En effet, si la nature exacte des *relata* causaux peut être discutée, aucun philosophe ne propose de considérer qu'ils sont des variables. Positivement, les conceptions les plus communément admises – auxquelles nous nous en tenons ici – font de la causalité générique une relation entre propriétés, et de la causalité singulière une relation entre événements.

On trouve dans Williamson (2005) une défense de l'idée selon laquelle les variables permettent de représenter d'une part des événements singuliers et d'autre part des propriétés :

> Il semble que ce soit une idéalisation sans conséquence que d'interpréter la causalité comme une relation entre variables – on peut penser un événement comme une variable binaire singulière qui prend une valeur s'il advient et l'autre valeur s'il n'advient pas, on peut penser une propriété comme une variable binaire répétable qui prend une valeur quand elle est instanciée et l'autre quand elle n'est pas instanciée, etc. – et une telle idéalisation entre rarement en conflit avec les intuitions causales.[24]

Nous discutons dans l'ordre les deux affirmations contenues dans cette citation.

Selon la première de ces affirmations, un événement singulier E peut être représenté par la variable binaire qui prend la valeur 1 (par exemple) si E ad-

[24] Williamson (2005) p. 50.

vient et la valeur 0 sinon – ce qui revient à considérer comme variable l'occurrence ou non de l'événement considéré. Cette proposition présente l'avantage d'énoncer une solution générale : pour tout événement possible, elle indique quelle variable binaire peut le représenter. C'est par ailleurs une réponse astucieuse. Elle consiste en effet à poser comme variable un paramètre parfaitement défini pour tout événement possible et qui, pour chacun de ces événements, prend deux valeurs qui suffisent à le définir et dont exactement une lui appartient. La variable non seulement représente, mais encore définit l'événement auquel elle renvoie. La proposition de Williamson paraît donc satisfaisante.

Toutefois, considérons maintenant non plus seulement des événements singuliers en général, mais deux événements singuliers tels que l'un cause l'autre – par exemple la dégustation de fruits de mer par tel de mes amis hier au soir et son mal de ventre dans la nuit de hier à aujourd'hui. La relation de cause à effet singulière que nous envisageons implique que ces deux événements sont advenus : mon ami a effectivement mangé des fruits de mer hier au soir et il a effectivement eu mal au ventre la nuit dernière. L'occurrence de l'un et/ou de l'autre de ces deux événements ne peut donc être sérieusement tenue pour variable. Si la proposition formulée par Williamson convient pour la représentation des événements singuliers possibles isolés, elle semble donc impropre pour la représentation de tels événements en tant qu'ils sont des *relata* causaux. On peut au mieux considérer que cette proposition indique comment les *relata* causaux peuvent être représentés par des valeurs de variables : l'événement singulier que constitue ma chute dans l'escalier hier matin à 8 heures peut être représenté par la valeur 1 d'une variable binaire qui prend la valeur 1 si et seulement si cet événement est advenu. Comme nous n'envisageons pas de mode de représentation des *relata* de la causalité singulière par des variables autre que celui que mentionne Williamson, nous considérons que les causes et effets singuliers ne peuvent pas être représentés par des variables. Dans ces conditions, la première composante de l'hypothèse de représentation implique que la notion de réseau bayésien causal n'a pas de sens si c'est de causalité singulière qu'il s'agit. Il convient maintenant d'expliquer comment et dans quelle mesure elle en a quand on parle de causalité générique.

Concernant le cas générique, nous souscrivons à la solution proposée par Williamson : « penser une propriété comme une variable binaire répétable qui prend une valeur quand elle est instanciée et l'autre quand elle n'est pas instanciée »[25]. De même que la proposition formulée par Williamson pour le cas singulier, cette proposition est générale au sens où pour toute propriété

[25]Williamson (2004) p. 50.

elle indique quelle variable binaire peut la représenter. Cette variable est plus précisément la variable qui prend la valeur 1 (par exemple) quand la propriété considérée est instanciée et la valeur 0 sinon. Contrairement à ce qui se passe dans le cas singulier, cette instanciation peut bien être considérée comme variable. Plus généralement, le déploiement de la proposition de Williamson fait apparaître que la représentation des propriétés au moyen de variables repose largement sur ceci que les valeurs de variables, comme les propriétés, sont (ou non) instanciées par les individus d'une population qu'on considère. Il apparaît par ailleurs que représenter une propriété au moyen d'une variable binaire peut être considéré comme un cas particulier limite de ce qu'une variable peut représenter.

Pour comprendre ce dernier point, arrêtons-nous à la propriété « être fumeur », qui entre dans des relations de cause à effet génériques bien connues et souvent discutées. La non-instanciation de cette propriété par un individu revient exactement à l'instanciation par ce même individu de la propriété « ne pas être fumeur ». L'ensemble de propriétés {être fumeur, ne pas être fumeur} est donc caractérisé par ceci que tout individu – et donc en particulier tout individu de la population qui sera considérée dans l'analyse – instancie exactement une des propriétés. En d'autres termes, l'ensemble {être fumeur, ne pas être fumeur} permet de définir une partition sur l'ensemble des individus considérés. Or, cette caractéristique n'appartient pas exclusivement à des ensembles de deux propriétés tels que l'une est définie comme la négation de l'autre. De nombreux ensembles de propriétés la possèdent, dont le cardinal peut avoir n'importe quelle valeur finie.[26] Parmi ces ensembles, on trouve en particulier des ensembles de propriétés dont la définition implique une valeur numérique – par exemple {avoir 20 ans ou moins, avoir entre 21 et 40 ans, avoir entre 41 et 60 ans, avoir 61 ans ou plus}. De même que les ensembles du type {être fumeur, ne pas être fumeur} et pour les mêmes raisons, ces ensembles sont représentés de manière adéquate par des variables. Ce sont d'ailleurs généralement des ensembles de ce type qui sont représentés par les variables des graphes bayésiens causaux.

L'analyse du paragraphe précédent a montré comment tout ensemble de propriétés tel que tout individu de la population considérée instancie exactement une de ces propriétés peut être représenté par une variable. Ce résultat, toutefois, peut être présenté sous un jour un peu différent. Il apparaîtra alors que nous avons montré que les variables ne représentent pas des propriétés, mais certains *ensembles de* propriétés. Dans ces conditions, une flèche

[26] On pourrait envisager également des ensembles de cardinal infini, seulement nous avons annoncé plus haut que notre intérêt se limitait dans ce texte aux ensembles finis de variables aléatoires discrètes étant chacune susceptible de prendre un ensemble fini de valeurs.

$V_C \longrightarrow V_E$ dans un graphe bayésien représente une relation entre *ensembles de* propriétés. Elle n'indique ni quelle(s) valeur(s) de V_C est (sont) une (des) cause(s), ni quelle(s) valeur(s) de V_E elle(s) cause(nt). A titre d'illustration, une flèche entre la variable binaire correspondant à la propriété d'être fumeur et la variable binaire correspondant à la propriété de développer un cancer n'indique pas si c'est fumer ou ne pas fumer qui cause le cancer. De façon similaire, une flèche entre une variable représentant la quantité d'eau qu'on donne à une plante et l'état de santé de cette plante n'indique pas quelle(s) quantité(s) d'eau cause(nt) quel(s) état(s) de santé de la plante. Si les *relata* de la causalité générique peuvent bien être représentés par des variables, la représentation des relations de cause à effet générique qui en découle est donc grossière.[27] Cette limite théorique a d'importantes conséquences pratiques : ainsi qu'il apparaît avec l'exemple précédent, ce qu'on perd d'information en représentant les *relata* de la causalité générique par des variables est un guide indispensable pour l'action.

En définitive, la première composante de l'hypothèse de représentation implique :

1. que la notion de réseau bayésien causal n'a de sens que relativement à la causalité *générique*. La première composante de l'hypothèse de représentation rend donc compte de ceci que les réseaux bayésiens causaux sont utilisés exclusivement dans des études portant sur la causalité générique, et non sur la causalité singulière ;
2. que la représentation de la causalité générique par les flèches d'un réseau bayésien causal est grossière et conduit à ne pas représenter certaines informations. De ces informations, on a vu pourtant qu'elles sont fondamentales d'un point de vue pratique.

La première composante de l'hypothèse de représentation a donc principalement des conséquences relatives à la granularité de la représentation de la causalité générique dans les graphes bayésiens causaux. L'analyse de cette première sous-hypothèse ayant étant menée à bien, nous nous tournons vers la deuxième composante de l'hypothèse de représentation.

1.2.2.2 Deuxième composante de l'hypothèse de représentation : Représenter la causalité directe par une relation binaire

Selon la deuxième composante de l'hypothèse de représentation, la relation de causalité directe au sein d'un ensemble de variables peut être

[27] Cette représentation est grossière en d'autres sens que celui que nous mettons explicitement au jour. En particulier, une flèche entre deux variables ne renseigne ni sur le mode d'action de la cause, ni sur les propriétés de la relation de cause à effet. Toutefois ces points ne nous intéressent pas directement ici et ne seront pas discutés dans ce chapitre.

représentée par une relation binaire. Cette deuxième sous-hypothèse correspond à ceci qu'une flèche (et donc en particulier une flèche d'un réseau bayésien) relie *deux* points. Elle semble peu discutable en première approche : une relation de cause à effet engage bien une cause d'une part et son effet de l'autre. Toutefois, elle apparaît plus problématique quand on la considère à la lumière des résultats de notre analyse de la première composante de l'hypothèse de représentation. En effet, certains auteurs, au premier rang desquels on trouve Eells[28], soutiennent que les relations de cause à effet génériques sont relatives à des populations. La causalité générique est alors une relation ternaire, dont les éléments sont de la forme (cause, effet, population).

La thèse selon laquelle la causalité générique est une relation ternaire est monnaie courante dans le domaine des théories probabilistes de la causalité. La raison en est la suivante : deux propriétés données peuvent entrer dans des rapports probabilistes[29] différents selon les populations qu'on considère. Nous aurons à revenir plus bas sur les cas de ce type, mais pour l'heure nous ne nous attardons pas à les classer et à les décrire. En effet, il suffit à notre argument présent d'avoir rendue plausible l'idée selon laquelle la relativisation des relations de cause à effet génériques à des populations est appelée par les approches probabilistes de la causalité. Puisque la notion de réseau bayésien causal est l'outil d'une telle approche, que cette notion soit corrélative d'une conception de la causalité générique comme relation binaire pose bien problème.

Pour résoudre ce problème, commençons par rappeler qu'un réseau bayésien est composé d'autre chose que du seul graphe orienté dont l'ensemble des flèches constitue une relation binaire sur l'ensemble des variables qui sont des sommets du graphe. Plus précisément, un réseau bayésien est un couple composé d'un graphe orienté et d'une distribution de probabilités. Cette distribution de probabilités est définie sur un ensemble de variables. D'après la définition 1.12 de l'appendice, il s'agit d'une fonction qui attribue une probabilité à toute valeur de l'ensemble de variables considéré. Si les valeurs que prend cette fonction dépendent de la population considérée, alors le réseau bayésien lui-même devient relatif à cette population et avec lui les couples ordonnés que le graphe bayésien représente.

En conséquence des résultats de notre analyse de la première composante de l'hypothèse de représentation, les probabilités attribuées aux différentes valeurs de l'ensemble de variables sur lequel un réseau bayésien est défini sont

[28] Eells (1991).

[29] Les rapports probabilistes que nous envisageons sont ceux qu'on trouve au fondement des caractérisations probabilistes de la causalité : augmentation de probabilité, diminution de probabilité, indépendance probabiliste.

des probabilités génériques. Elles peuvent être en particulier des fréquences relatives – dans des suites finies ou infinies, réelles ou hypothétiques – et c'est ce qu'elles sont le plus souvent dans les utilisations effectives des réseaux bayésiens. Dans ce cas, elles sont bien relatives à une population, celle des individus appartenant à la suite relativement à laquelle les fréquences sont définies. Mais même si ce n'est pas le cas, c'est-à-dire même si les probabilités génériques ne sont pas interprétées comme des fréquences relatives, il reste que le réseau bayésien est relatif à une population dans la mesure où la distribution de probabilités qui le compose l'est elle-même. Ici comme dans les théories probabilistes de la causalité, l'arité de la relation de causalité dépend de la façon dont sont définies les probabilités.

De façon plus générale, nous avons montré que replacer les graphes bayésiens au sein des *couples* que sont les réseaux bayésiens permet de rejeter l'idée selon laquelle la notion de réseau bayésien causal interdirait de penser la causalité générique comme une relation ternaire composé de triplets (cause, effet, population). La relation de causalité représentée par les flèches d'un graphe bayésien est relative à une population si la distribution de probabilités du réseau bayésien considéré l'est elle-même. Les questions qui restent alors (e.g. faut-il effectivement considérer la causalité générique comme une relation ternaire ? si oui, quelles sont les populations à considérer ? comment les reconnaître ?) se posent classiquement dans le contexte d'approches probabilistes de la causalité. Le fait qu'elles soient soulevées par la notion de réseau bayésien causal ne pose pas de difficultés qui seraient spécifiques. Dans ces conditions, la deuxième composante de l'hypothèse de représentation ne doit pas être considérée comme substantielle. Par commodité, nous parlerons dans la suite comme si la relation de causalité générique était binaire, mais tout ce que nous dirons pourra être transcrit selon les voies indiquées ici en affirmations relatives à la causalité générique considérée comme une relation ternaire.

1.2.2.3 Troisième composante de l'hypothèse de représentation : Représenter la causalité par une relation asymétrique

La troisième composante de l'hypothèse de représentation prolonge la deuxième. En effet, de cette relation de causalité directe qui selon la sous-hypothèse 2 représente les relations de cause à effet directes au sein d'un ensemble de variables, la sous-hypothèse 3 énonce une propriété : l'asymétrie. L'asymétrie de la relation binaire qui représente la causalité directe dans un réseau bayésien causal découle de ceci que les graphes bayésiens sont orientés et acycliques.

La troisième composante de l'hypothèse de représentation semble peu

problématique dans la mesure où la causalité elle-même se présente comme une relation asymétrique. Toutefois, ce ne sont pas les relations de cause à effet en général, mais les relations de cause à effet *directes*, au sein de l'ensemble de variables sur lequel il est défini que représente le graphe d'un réseau bayésien causal. Il convient donc de rendre compte précis du rapport entre le caractère asymétrique de la causalité d'une part et d'autre part l'acyclicité du graphe qui représente la causalité *directe* parmi les variables de l'ensemble sur lequel un réseau bayésien est défini.

Pour mener à bien cette première tâche, revenons un moment sur la notion de relation de cause à effet *directe*. Nous avons vu que le caractère direct ou non d'une relation de cause à effet dépend de l'ensemble de variables considéré. Il est ainsi apparu que la variable G correspondant à la propriété d'être porteur d'un gène de susceptibilité pour le cancer du sein peut être une cause directe de la variable C correspondant à la propriété de recevoir une chimiothérapie, mais qu'elle n'en est pas une relativement à un ensemble auquel appartient la variable A correspondant la propriété d'être atteint d'un cancer. Pourtant, même dans le cas où A appartient à l'ensemble de variables considéré, G est une cause (certes non directe) de C. La causalité, en effet, se présente comme une relation transitive. Plus exactement, nous définissons la causalité sur un ensemble de variables comme la clôture transitive de la causalité directe sur cet ensemble.

Sous cette définition de la causalité, on peut montrer que l'acyclicité du graphe qui représente la causalité directe est équivalente à l'asymétrie de la causalité.

Preuve : Soit un graphe orienté G représentant les relations de cause à effet directes sur un ensemble de variables $\mathbf{V}$. Les propositions suivantes sont alors équivalentes :

- la causalité sur $\mathbf{V}$ n'est pas une relation asymétrique ;
- il existe deux variables V_i et V_j de $\mathbf{V}$ telles que chacune cause l'autre ;
- G contient un sous-graphe de la forme suivante : $V_i \leftrightarrows \cdots \leftrightarrows V_j$;
- G n'est pas acyclique.

Nous venons de montrer que l'hypothèse selon laquelle le graphe qui représente les relations de cause à effet directes sur un ensemble de variables est acyclique est équivalente à l'affirmation selon laquelle la causalité est une relation asymétrique. Or, la causalité se présente effectivement comme une relation asymétrique. Un effet ne cause pas sa cause. Dans ces conditions, la troisième composante de l'hypothèse de représentation ne soulève pas de difficultés particulières. Pour le dire autrement, le bien-fondé de la troisième composante de l'hypothèse de représentation est assuré par les propriétés même de la causalité.

Nous venons d'analyser l'hypothèse de représentation, selon laquelle les relations de cause à effet directes sur un ensemble de variables peuvent être représentées par des flèches du type de celles qui figurent dans le graphe d'un réseau bayésien défini sur $\mathbf{V}$, et de discuter les trois sous-hypothèses qui composent cette hypothèse. La discussion a fait apparaître que :

1. la première composante de l'hypothèse de représentation n'est pas critiquable puisqu'elle doit être comprise essentiellement comme une convention relative au grain de la représentation des relations de cause à effet génériques directes ;
2. la deuxième composante de l'hypothèse de représentation n'est pas substantielle. En effet nous avons vu que, dans le contexte d'approches probabilistes de la causalité, l'arité de la relation de causalité dépend de la façon dont les probabilités sont définies. Dès lors, la représentation des relations de cause à effet au moyen des flèches qui relient *deux* variables est compatible à la fois avec une notion binaire, et avec une notion ternaire, de causalité ;
3. la troisième composante de l'hypothèse de représentation est substantielle au sens où, contrairement aux deux précédentes, elle véhicule une hypothèse sur la causalité elle-même. Cette hypothèse est exactement celle de l'asymétrie de la causalité considérée comme clôture transitive de la causalité directe. Dans la mesure où la causalité est effectivement asymétrique, nous avons conclu au caractère non problématique de la troisième composante de l'hypothèse de représentation.

Dans cette sous-section, nous nous sommes intéressés aux corrélats de la notion de réseau bayésien causal relativement à la causalité elle-même. Or, nous l'avons déjà indiqué, la notion de réseau bayésien causal a également des corrélats relatifs au rapport entre causalité et probabilités. C'est à eux qu'il convient que nous en venions maintenant, sous la forme d'une discussion de la condition de Markov causale.

1.2.3 La condition de Markov causale

Ainsi que nous l'avons annoncé en préambule à la présente section, la condition de Markov causale est l'hypothèse selon laquelle le graphe qui représente les relations de cause à effet directes entre les variables d'un ensemble est compatible avec la distribution de probabilités sur cet ensemble. On peut, maintenant, expliciter cette hypothèse. Pour cela, rappelons d'abord que, en vertu de la proposition 1.1 établie par Pearl, les réseaux bayésiens peuvent être définis non seulement par la notion de représentation d'une distribution de probabilités par un graphe orienté acyclique (définition

1.4), mais encore par la condition de Markov. Rappelons ensuite que la condition de Markov est satisfaite par un couple (G, p) sur $\mathbf{V}$ si et seulement si toute variable de $\mathbf{V}$ est indépendante pour p de tous ses non-descendants dans G relativement à l'ensemble de ses parents dans $\mathbf{V}$. Il en découle que la notion de réseau bayésien causal suppose exactement ceci relativement au rapport entre causalité et probabilités :

Définition 1.9 (Condition de Markov causale (CMC)) *Etant donné un ensemble de variables $\boldsymbol{V}$, toute variable de $\boldsymbol{V}$ est indépendante de tous ses non-descendants causaux relativement à l'ensemble de ses causes directes dans $\boldsymbol{V}$.*

Dans la sous-section qui commence, nous analysons la condition de Markov causale et montrons que la causalité et les probabilités semblent bien entretenir des rapports du type de ceux qu'elle décrit. Préalablement à cette analyse et à cette discussion, il convient toutefois de s'arrêter un instant à la question de la nature de la distribution de probabilités – ou, en d'autres termes, à l'interprétation des probabilités – dont il est question dans la condition de Markov causale.

1.2.3.1 Nature des probabilités

Remarquons, pour commencer, que la question de l'interprétation à donner aux distributions de probabilités qui apparaissent dans les réseaux bayésiens causaux n'est presque jamais traitée pour elle-même. Une exception notable est Williamson (2005), dont l'objet principal est de déterminer ce que sont ou peuvent être les relations de cause à effet et les probabilités qui figurent dans les réseaux bayésiens causaux. Mais, précisément, ce texte traite trop spécifiquement de cette question et insuffisamment des usages effectifs des réseaux bayésiens causaux pour nous être véritablement utile ici. Ce qui nous intéresse dans ce chapitre, et plus généralement dans toute la première partie de notre travail, est l'utilisation des réseaux bayésiens causaux dans des contextes d'inférence aux causes à partir de prémisses probabilistes (second des deux types d'utilisations des réseaux bayésiens causaux mentionnés plus haut). Or, il nous semble clair que, dans ce contexte, les probabilités sont des degrés de croyance : qui tire une inférence dont il prétend la conclusion vraie, prend pour prémisses des propositions qu'il croit vraies. Cela n'implique pas, bien entendu, que les probabilités qui nous intéressent ici requièrent une interprétation subjectiviste radicale, du type de celle que propose de Finetti par exemple. Pour dire les choses autrement, nous laissons ouverte la question de savoir comment ces degrés de croyance sont formés, quel type de contraintes il convient de faire peser sur l'évaluation des probabilités.

Reste, maintenant, que la question de la vérité de la condition de Markov causale devient difficile à envisager sous une interprétation subjective des probabilités. Plus précisément, nous voyons mal quel sens il y a à demander si mes degrés de croyance subjectifs sont compatibles avec le graphe qui représente les relations de cause à effet directes sur cet ensemble. Il apparaît alors que la question de la validité de la condition de Markov causale n'est intelligible que pour une interprétation des probabilités telle que la valeur des probabilités ne varie pas avec les individus.

Williamson qualifie d'« objectives »[30] les probabilités dont l'évaluation ne varie pas avec les individus. Il insiste sur ceci que la distinction physique / épistémique ne coïncide pas, en philosophie des probabilités, avec la distinction objectif / subjectif. Positivement, il développe une interprétation sous laquelle les probabilités sont des degrés de croyance, mais des degrés de croyance sur la formation desquels pèsent de telles contraintes – logiques et empiriques – qu'ils sont objectifs au sens que nous venons d'introduire. Cette interprétation est connue sous le nom de « bayésianisme objectif ».

Pour deux raisons, nous ne posons pas la question de la vérité de la condition de Markov causale pour des probabilités interprétées dans les termes du bayésianisme objectif. En premier lieu, la question ainsi posée est déjà résolue par Williamson – du moins si la causalité est épistémique au sens où il définit ce terme. En second lieu, et surtout, ce qui nous intéresse dans ce travail est ce qu'on peut inférer des probabilités telles que nous les évaluons ordinairement, et plus précisément ce qu'on peut en inférer relativement à ce qu'on entend habituellement par « causalité ». Nous poserons donc la question comme une question relative au rapport entre ce qu'on reconnaît habituellement comme les probabilités et ce qu'on considère usuellement être la causalité. Autrement dit, nous poserons la question de la vérité de la condition de Markov causale indépendamment de toute interprétation précise des probabilités et de toute théorie articulée de la causalité. Cette position pourrait paraître faible, mais on notera qu'elle est tacite dans toute la littérature relative à la condition de Markov causale en particulier, et aux réseaux bayésiens causaux en général : à l'exception notable de Williamson (2005), la question de l'interprétation des probabilités n'y est presque jamais discutée. Finalement, on soulignera que la position que nous adoptons n'implique pas que nous renonçons à la thèse selon laquelle, dans le contexte d'inférence causale, les probabilités qui apparaissent dans les réseaux bayésiens causaux ne peuvent être que des degrés de croyance.

Maintenant que nous avons explicité les termes dans lesquels la discussion sera menée, il convient d'analyser la condition de Markov.

[30] Williamson (2005) p. 7.

1.2.3.2 Analyse de la condition de Markov causale

Selon le théorème 1.1 énoncé plus haut, la condition de Markov causale implique que deux ensembles de variables d-séparés dans le graphe causal G sur $\mathbf{V}$ sont indépendants (en probabilités) relativement au sous-ensemble de $\mathbf{V}$ qui les d-séparent dans G. Dès lors, il semble clair que la condition de Markov causale est une hypothèse selon laquelle la structure causale fait peser un grand nombre de contraintes sur les probabilités – ou plus exactement sur la structure des indépendances probabilistes. Or, nous ne voyons pas comment construire une typologie exhaustive des cas d'indépendances probabilistes imposées par la structure causale selon la condition de Markov causale. Dans ces conditions, la stratégie que nous adoptons pour analyser et discuter la condition de Markov causale est la suivante : nous nous concentrons sur les indépendances probabilistes auxquelles elle fait explicitement référence, nous identifions trois grands types de telles indépendances, que nous discutons tour à tour.

Etant donné un ensemble de variables $\mathbf{V}$, les indépendances probabilistes explicitement mentionnées par la condition de Markov causale sont les indépendances d'une variable V_1 de $\mathbf{V}$ et d'une variable V_2 de $\mathbf{V}$ relativement à l'ensemble des causes directes de V_1 dans $\mathbf{V}$. Plus précisément, selon la condition de Markov causale, les variables dont une variable V_1 est indépendante relativement à l'ensemble de ses causes directes dans $\mathbf{V}$ sont ses non-descendants dans le graphe causal sur $\mathbf{V}$. Or, un non-descendant de V_1 dans le graphe causal sur $\mathbf{V}$ peut être :

0. une cause directe de V_1 dans $\mathbf{V}$;
1. une variable qui n'a aucune sorte de rapport causal avec V_1 ;
2. un antécédent causal de V_1 qui n'en est pas une cause directe dans $\mathbf{V}$;
3. un descendant causal d'une cause directe de V_1 dans $\mathbf{V}$ qui n'est pas un descendant causal de V_1.

Selon la condition de Markov causale, V_1 est indépendante relativement à l'ensemble de ses causes directes de toute variable V_2 relevant des quatre types que nous venons d'énumérer. On dira que l'ensemble des causes directes de V_1 « fait écran » entre V_1 et chacune des variables relevant de l'un des quatre types identifiés. Pour les variables du premier type, l'indépendance est analytique au sens où elle découle des propriétés des probabilités conditionnelles. En effet, on montre facilement la proposition suivante :

Proposition 1.3 *Soit p une distribution de probabilités sur un ensemble de variables $\boldsymbol{V}$.*
Pour tout $\boldsymbol{W} \subset \boldsymbol{V}$, tout $W \in \boldsymbol{W}$ et tout $V \in \boldsymbol{V}$, V est indépendant de W relativement à $\boldsymbol{W}$.

Pour les variables des types 1. à 3., l'indépendance n'est pas une conséquence des propriétés des probabilités conditionnelles. En ce sens, l'hypothèse de l'indépendance est substantielle dans les cas relevant de ces trois types. Dès lors, nous envisageons chacune à son tour les trois sous-hypothèses de la condition de Markov qui se font jour ici. Pour chacune, nous nous attachons en particulier à montrer qu'elle ne semble pas problématique, c'est-à-dire qu'elle semble être effectivement une propriété du rapport entre la structure causale et les probabilités.

1.2.3.3 Sous-hypothèse 1. : Probabilités et indépendance causale

Selon la première sous-hypothèse de la condition de Markov que nous avons mise au jour, une variable V_1 de $\mathbf{V}$ est indépendante relativement à l'ensemble de ses causes directes dans $\mathbf{V}$ de chacune des variables de $\mathbf{V}$ avec lesquelles elle n'entretient aucune forme de relation causale. En termes graphiques, V_1 est indépendante de toute variable V_2 telle qu'il n'existe pas de chemin (orienté ou non) entre V_1 et V_2 dans le graphe qui représente la causalité directe sur $\mathbf{V}$. Par contraposition, il apparaît que cette première sous-hypothèse de la condition de Markov causale peut s'énoncer de la façon suivante : toutes les dépendances probabilistes renvoient à des dépendances causales.

Ainsi reformulée, cette première hypothèse se présente comme une exigence d'explication[31] des variations, et plus précisément des co-variations. Cette exigence est une exigence scientifique classique, dont Mill propose une formulation explicite :

> Quelque phénomène que ce soit qui varie de quelque façon que ce soit quand un autre phénomène varie d'une certaine façon soit est une cause ou un effet de ce phénomène, soit est connecté avec lui par quelque fait causal.[32]

Dans la mesure où elle se présente comme inséparable de l'exigence scientifique, la première sous-hypothèse de la condition de Markov causale semble ne pas être problématique, mais au contraire décrire adéquatement un aspect du rapport entre causalité et probabilités. Nous en venons donc à la deuxième sous-hypothèse de la condition de Markov causale.

[31] La notion d'explication est utilisée ici dans un sens peu précis mais qui nous semble suffisamment clair. Cette notion, et en particulier les rapports précis qu'elle entretient avec celle de causalité, n'est pas l'objet de notre analyse.

[32] Mill (1843) p. 287. Cette référence est reprise de Williamson (Williamson (2005) p. 51).

1.2.3.4 Sous-hypothèse 2. : Probabilités et causalité indirecte, contiguïté des causes et de leurs effets

En deuxième lieu la condition de Markov causale implique qu'une variable V_1 de $\mathbf{V}$ est indépendante, relativement à l'ensemble de ses causes directes dans $\mathbf{V}$, de ses autres antécédents causaux. Sous la condition de Markov causale, l'influence des antécédents causaux se résume à celle des antécédents causaux immédiats. Ainsi reformulée, la deuxième sous-hypothèse de la condition de Markov causale apparaît comme sa partie proprement markovienne : seules les causes directes sont pertinentes pour l'état d'une variable – ici : pour la distribution de probabilités sur cette variable.

Cette partie proprement markovienne de la condition de Markov causale décrit une propriété qui semble bien appartenir au rapport entre la causalité et les probabilités. Plus précisément, nous défendons qu'elle est une version probabiliste de la thèse selon laquelle une cause et son effet sont contigus. La thèse de la contiguïté des causes et de leurs effets se présente comme une évidence intuitive : il semble clair qu'une action causale n'est que d'un proche à un proche. Pour son caractère d'évidence intuitive, Hume fait entrer la thèse de la contiguïté des causes et des effets au nombre des affirmations qui caractérisent essentiellement la causalité :

> tous les objets qu'on considère comme causes et comme effets sont *contigus*.[33]

Plus récemment, la contiguïté des causes et de leurs effets est celle des caractéristiques classiques de la causalité dont rendent compte au premier chef les analyses de la causalité comme transfert. Selon ces théories, ce qui fait que A cause B est ceci que quelque chose est transférée de A à B. Cette thèse générale se spécifie chez Reichenbach[34] et Salmon[35], et devient l'idée selon laquelle les processus causaux se caractérisent par leur capacité à transmettre une marque. Elle a été défendue ensuite comme thèse de la conservation dans l'effet d'une quantité présente dans la cause.[36]

Au-delà de l'attrait que la thèse de la contiguïté des causes et des effets exerce manifestement sur les théoriciens de la causalité, nous retenons que cette thèse est étayée par des résultats physiques :

> Aujourd'hui, nous pouvons compter sur le résultat scientifique selon lequel toutes les forces fondamentales n'agissent à distance que grâce à une propagation préalable qui, elle, a une vitesse finie. [...] En ce qui

[33] Hume (1739) p. 134.
[34] Reichenbach (1956) chapitre 23.
[35] Salmon (1984).
[36] Aronson (1971), Fair (1979), Dowe (1992a, 1992b), Kistler (1999).

> concerne le mécanisme de l'action à distance des forces fondamentales, la physique nous donne des arguments solides pour soutenir qu'elle est le résultat de processus causaux sous-jacents qui obéissent strictement à la contiguïté.[37]

Nous n'entrons dans l'analyse des points mentionnés par Kistler, mais concluons de ce passage que la thèse de la contiguïté des causes et de leurs effets n'est pas seulement attirante d'un point de vue intuitif. De même que Max Kistler un peu plus loin dans le texte[38], nous considérerons donc que cette thèse est vraie.

La vérité de la thèse de la contiguïté des causes et de leurs effets n'établit pas directement que la deuxième des sous-hypothèses qui constituent la condition de Markov causale est elle-même vraie. La raison principale en est que les causes et les effets dont nous venons de voir qu'ils sont contigus – et avec eux la causalité qu'on analyse comme transfert ou comme processus – sont *singuliers*. La thèse de la contiguïté n'a d'ailleurs pas de sens dans le cas générique puisqu'elle suppose, pour être intelligible, que les *relata* causaux soient situés dans l'espace et dans le temps. Or, ainsi que nous l'avons montré à l'occasion de l'analyse de la première composante de l'hypothèse de représentation, les relations de cause à effet que peuvent représenter les flèches des réseaux bayésiens sont *génériques*. En outre, la thèse de la contiguïté des causes et de leurs effets ne fait pas explicitement référence à des probabilités, alors que la deuxième sous-hypothèse de la condition de Markov causale porte sur des indépendances probabilistes. On comble d'un même mouvement les deux fossés que nous venons de mettre en évidence en considérant que la thèse de la contiguïté des causes et de leurs effets est l'expression singulière de la thèse plus générale selon laquelle les influences causales se propagent de proche en proche.

La thèse de la propagation de proche en proche est la thèse de la contiguïté des causes et de leurs effets en tant qu'elle est étendue du niveau singulier auquel elle trouve son sens premier, spatio-temporel, au niveau générique qui nous intéresse dans cette première partie de notre travail. En effet, et en premier lieu, cette thèse plus générale peut trouver une expression relative à la causalité générique qui est représentée par les réseaux bayésiens causaux. En second lieu, l'influence causale se comprend dans ce contexte justement et exactement comme dépendance probabiliste. Dans ces conditions, dire que les influences causales se propagent de proche en proche c'est dire en particulier – et concernant précisément les réseaux bayésiens causaux – que la dépendance probabiliste qu'une variable entretient à l'égard de ses causes

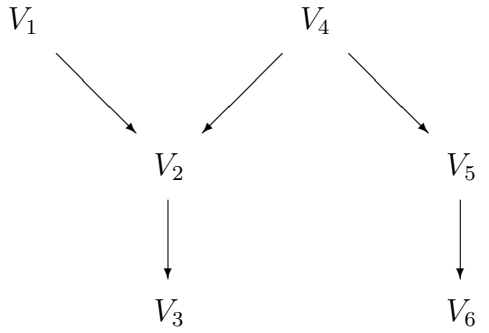
[37]Kistler (1999) pp. 40–41.
[38]Kistler (1999) p. 43.

est tout entière contenue dans la dépendance qu'elle entretient à l'égard de ses causes directes. Il en découle que conditionnaliser sur ces causes directes rend la variable considérée indépendante de toutes ses autres causes. C'est là exactement la deuxième sous-hypothèse de la condition de Markov causale.

Nous avons montré deux choses distinctes dans ce paragraphe : d'une part que la thèse de la contiguïté des causes singulières et de leurs effets peut être acceptée, d'autre part que cette thèse et la deuxième sous-hypothèse de la condition de Markov causale sont deux expressions de la thèse plus générale selon laquelle l'influence causale se propage de proche en proche. Bien sûr, la conjonction de ces deux résultats ne constitue pas un argument définitif en faveur de la vérité de la deuxième sous-hypothèse de la condition de Markov causale. Toutefois il nous semble qu'elle contribue de façon significative à rendre cette hypothèse plausible. En effet, la propagation des influences causales n'existe physiquement que comme propagation des influences causales singulières et on voit dès lors mal comment la valeur de vérité de la thèse de la contiguïté pourrait différer de celle la thèse plus générale selon laquelle les influences causales – et en particulier les influences causales génériques représentées dans les réseaux bayésiens causaux – se propagent de proche en proche. Dans ces conditions, il nous semble avoir au moins établi qu'il est plausible que la deuxième sous-hypothèse de la condition de Markov causale décrit adéquatement un aspect du rapport entre causalité et probabilités. Il nous reste à faire le même travail pour la troisième sous-hypothèse de la condition de Markov causale.

1.2.3.5 Sous-hypothèse 3. : Probabilités et causes communes

Selon la troisième des sous-hypothèses qui composent la condition de Markov causale, une variable V est indépendante relativement à l'ensemble de ses causes directes dans $\mathbf{V}$ de toute variable de $\mathbf{V}$ qui est un effet de l'une de ces causes mais pas de V. A titre d'illustration, considérons le graphe suivant :



Dans ce graphe, la troisième sous-hypothèse de la condition de Markov implique que :
V_2 est indépendante de V_5 et V_6 relativement à $\{V_1, V_4\}$;
V_5 est indépendante de V_2 et de V_3 relativement à V_4.
Il apparaît alors que la troisième sous-hypothèse de la condition de Markov causale peut être considéré comme une version du principe de la cause commune.

Le principe de la cause commune est clairement identifié et discuté par Reichenbach.[39]. Concernant des variables, il peut être formulé de la façon suivante :

Définition 1.10 (Principe de la cause commune (PCC)) *Si deux variables V_1 et V_2 sont dépendantes et qu'aucune des deux n'est une cause de l'autre, alors il existe une variable V_3 qui est une cause de V_1 et une cause de V_2 et qui est telle que V_1 et V_2 sont indépendantes relativement à elle.*

Cette formulation fait apparaître des différences entre le principe de la cause commune et la troisième sous-hypothèse de la condition de Markov causale. Nous discutons ces différences – et, positivement, rendons compte du rapport entre les deux énoncés qui nous intéressent ici – en deux temps.

En premier lieu, nous remarquons que la troisième sous-hypothèse de la condition de Markov causale fait référence à des causes directes, là où le principe de la cause commune traite de causalité en général, non spécifiquement directe. Cette première différence se réduit à partir de la remarque suivante : dans un réseau bayésien causal, s'il existe V_3 qui est une cause commune à V_1 et V_2 et une cause directe de V_1, alors qu'il existe une cause commune à V_1 et V_2 qui les rend indépendantes implique que V_3 les rend indépendantes. Cette propriété résulte de ce que les influences causales se propagent de proche en proche dans un réseau bayésien, où plus exactement de la deuxième composante de la condition de Markov causale. De cette propriété, il découle que dans les réseaux bayésiens causaux le principe de la cause commune a la conséquence suivante : étant données une variable V_1, V_3 une cause directe de V_1 et V_2 un effet de V_3 qui n'est pas causé par V_1, V_1 et V_2 sont indépendantes relativement à V_3. Autrement dit, le principe de la cause commune implique l'indépendance relativement à une cause commune qui est une cause directe de l'une des variables considérées.

Mais ce n'est pas là exactement la troisième sous-hypothèse de la condition de Markov causale. En second lieu, en effet, cette sous-hypothèse fait référence à la conditionnalisation sur *l'ensemble* des causes directes d'une variable donnée, là où le principe de la cause commune ne mentionne que

[39]Reichenbach (1956)pp. 158–159.

la conditionnalisation sur *une* cause. Ainsi que le note Arntzenius, prendre en compte l'ensemble des causes plutôt que l'une d'elles seulement est une modification du principe de la cause commune qui

- « clairement ne viole pas l'esprit du principe de la cause commune de Reichenbach »[40] ;
- étend le domaine de validité du principe. En effet, il existe des couples d'effets ayant au moins deux causes communes, qui ne sont indépendants relativement à aucun de ces causes communes prise séparément, mais qui sont indépendants relativement à l'ensemble de ces causes communes.

Dans ces conditions, cette seconde différence entre le principe de la cause commune et la troisième sous-hypothèse de la condition de Markov causale marque une supériorité de celle-ci sur celui-là.

Nous avons montré d'abord que le principe de la cause commune implique, dans le contexte que constituent les réseaux bayésiens causaux, des indépendances qui relèvent de la troisième sous-hypothèse de la condition de Markov causale – précisément les indépendances entre deux effets relativement à une cause commune qui est une cause directe de l'un des effets considérés. Dans ces conditions, ce type particulier d'indépendances impliquées par la troisième sous-hypothèse de la condition de Markov causale hérite de la plausibilité et de l'utilité généralement reconnues au principe de la cause commune. Nous avons montré ensuite que la différence entre les indépendances impliquées par la troisième sous-hypothèse de la condition de Markov causale et celles de ces indépendances qui relèvent du principe de la cause commune est telle que les premières sont plus souvent le cas que les secondes. Notre conclusion en sort renforcée : la troisième sous-hypothèse de la condition de Markov causale semble bien décrire adéquatement un aspect du rapport entre la causalité et les probabilités.

Pour en finir avec le traitement de la troisième sous-hypothèse de la condition de Markov causale, il convient de mentionner le rapport suivant entre la condition de Markov causale et le principe de la cause commune :

Proposition 1.4 (CMC et PCC) *La condition de Markov causale implique le principe de la cause commune.*

La preuve est simple ; on peut se référer à celle que Williamson donne.[41] En outre, cette proposition appelle deux remarques. Selon la première, la réciproque de l'implication qu'elle énonce n'est pas vraie. Ainsi, la condition de Markov causale n'est pas satisfaite dès lors que l'est le principe de la cause

[40] Arntzenius (2005) section 1.1.
[41] Williamson (2005) pp. 51–52.

commune. A titre d'illustration, on peut considérer un ensemble $\{V_1, V_2, V_3\}$ de trois variables, tel que le graphe représentant les relations de cause à effet directes est $V_1 \longrightarrow V_2 \longrightarrow V_3$. Si V_1 n'est pas indépendante de V_3 relativement à V_2, alors le principe de la cause commune est (trivialement) valide alors que la condition de Markov causale est violée. Notre seconde remarque concernant la proposition 1.4 est la suivante : elle n'est pas mise au premier plan par la décomposition que nous avons adoptée pour la condition de Markov causale. Nous nous tenons toutefois à cette décomposition en raison de son caractère littéral et de la valeur de clarification que nous lui reconnaissons. En mentionnant néanmoins la proposition 1.4, nous nous donnons la possibilité de revenir plus bas sur ce point.

Dans la section qui s'achève, nous avons analysé et discuté les hypothèses corrélatives de la notion de réseau bayésien causal. La réalisation de cette tâche a été guidée par la distinction entre les hypothèses portant sur la causalité elle-même et les hypothèses portant sur le rapport entre la causalité et les probabilités. Mais l'ensemble de la discussion fait apparaître que cette typologie n'est pas la plus pertinente. Nous distinguerons maintenant entre :

1. une hypothèse faisant état de la granularité de la représentation de la causalité dans les réseaux bayésiens causaux. Il s'agit de la première composante de l'hypothèse de représentation, et elle implique que la représentation des relations de cause à effet génériques dans les réseaux bayésiens causaux est plutôt grossière ;
2. deux hypothèses portant sur les relations causales et non sur leur représentation. Plus exactement, il s'agit de deux hypothèses dont on ne voit pas comment elles pourraient être satisfaites sans que la causalité elle-même ait effectivement certaines propriétés. Selon la première de ces hypothèses, la causalité générique est asymétrique. Selon la seconde, graphes causaux et distributions de probabilités sont dans le rapport énoncé par la condition de Markov. Nous avons mis au jour trois types fondamentaux d'indépendances impliquées par la condition de Markov devenue causale. Pour chacun, nous nous sommes attachés à rendre plausible l'hypothèse selon laquelle les indépendances concernées valent effectivement. De la même façon, nous avions fait état de la plausibilité de la thèse selon laquelle la causalité générique est asymétrique.

Si cette typologie laisse de côté la deuxième composante de l'hypothèse de représentation, c'est que nous avons établi qu'elle n'est pas substantielle, au sens où elle n'a pas de conséquence si on la considère indépendamment de la définition des distributions de probabilités qui composent les réseaux bayésiens causaux.

En définitive, nous pensons avoir rendu compte du bien-fondé de la notion de réseau bayésien causal. Il reste toutefois que les hypothèses de type 2., si elles apparaissent plausibles et valent dans de nombreux cas, se heurtent à des contre-exemples. Ce sont ces contre-exemples, ainsi que les artifices qui ont été envisagés pour les contourner, que nous présentons et discutons maintenant.

1.3 Contre-exemples aux hypothèses relatives à la causalité

Dans cette section, nous revenons d'abord sur la troisième composante de l'hypothèse de représentation et ensuite sur la condition de Markov causale. Pour chacune, nous commençons par montrer qu'elle admet des contre-exemples, puis nous expliquons comment les cas problématiques peuvent être traités dans le format des réseaux bayésiens causaux moyennant certains artifices. Une dernière sous-section est consacrée à tracer les limites du domaine dans lequel ces artifices peuvent être effectivement utilisés.

1.3.1 Acyclicité de la causalité générique

Selon la troisième composante de l'hypothèse de représentation, le graphe qui représente les relations de cause à effet directes sur un ensemble de variables est acyclique. Nous avons montré que cette hypothèse est équivalente à l'asymétrie de la causalité définie comme la clôture transitive de la relation de causalité directe. L'hypothèse est dès lors apparue peu problématique : la causalité, en effet, se présente bien comme une relation asymétrique. Pour le dire autrement, un effet ne cause jamais sa cause.

1.3.1.1 Cycles causaux

L'argument le plus fort en faveur de la thèse selon laquelle la causalité est asymétrique est celui qui consiste à faire valoir que les causes précèdent temporellement leurs effets. Cette précédence temporelle peut être considérée comme un élément de la définition de la causalité ou comme une propriété, éventuellement contingente, de toutes les relations de cause à effet que nous connaissons ; dans tous les cas, elle implique bien l'asymétrie de la relation de causalité. Cet argument, toutefois, vaut dans le seul cas singulier. Plus précisément, il n'y a pas de sens à dire qu'une propriété C est antérieure à une autre propriété E, et donc *a fortiori* à dire qu'une variable V_1 est antérieure à une variable V_2. En ce qui concerne les propriétés, seules leurs

instanciations par des individus singuliers sont temporellement ordonnées ; en ce qui concerne les variables, seules le sont les instanciations par des individus singuliers de propriétés appartenant à la famille qu'elles représentent. Autrement dit, notre meilleur argument en faveur de l'asymétrie de la causalité ne vaut pas pour la relation de causalité générique qui nous intéresse dans cette première partie de notre travail. Plus spécifiquement, il ne vaut pas pour la relation de causalité entre variables qui est représentée dans les réseaux bayésiens causaux. L'argument temporel ne peut donc pas être mobilisé pour montrer que la troisième composante de l'hypothèse de représentation n'est pas problématique, au sens où elle ne supposerait rien qui ne soit effectivement vrai.

Nous venons de voir que l'argument temporel en faveur de l'asymétrie de la causalité ne vaut pas dans le cas générique. Il y a, cependant, pire que cela : à y regarder de plus près, il semble possible que la relation de causalité générique ne soit pas, finalement, asymétrique. Ainsi Williamson formule-t-il la remarque suivante :

> Les cycles causaux sont en fait répandus : la pauvreté cause le crime qui cause plus de pauvreté ; un système immunitaire faible conduit à la maladie qui peut encore affaiblir le système immunitaire ; les augmentations des prix de l'immobilier cause un empressement à acheter qui en retour cause de nouvelles augmentations des prix.[42]

Certains auteurs maintiennent la thèse de l'asymétrie de la causalité générique en dépit de contre-exemples de ce type. Ces auteurs doivent principalement montrer que ce n'est qu'en apparence que la causalité générique n'est pas asymétrique dans les cas du type de ceux que Williamson mentionne. Eells est celui qui entreprend cette tâche avec le plus de rigueur et pour le résultat le plus abouti[43] ; c'est donc son analyse qui retient notre attention dans la fin de ce paragraphe. Or, cette analyse suppose de considérer 1) que tout individu a est un ensemble de tranches temporelles de substance individuelle : $a = \{a_t\}$ où t est un instant de la durée de a, 2) qu'il existe des propriétés indexées sur le temps, définies de la façon suivante : pour tout F, tout t, tout a, $F_t(a) =_{def} F(a_t)$, 3) que les énoncés causaux singuliers font référence à de telles propriétés et 4) que les affirmations causales génériques sont relatives à des distances temporelles et s'analysent comme des quantifications universelles , sur l'ensemble des couples ordonnés d'instants temporels adéquatement distants, d'énoncés causaux singuliers. Chacune de ces quatre propositions rompt avec certaines de nos intuitions et impliquent une distance entre la façon dont parlons des choses et ce qu'elles sont réellement. Accepter

[42] Williamson (2005) p. 50. Davis formule une remarque du même type (Davis, 1988, p. 146.)

[43] Eells (1991) pp. 48 et suivantes.

ces quatre propositions ensemble nous semble être payer un prix trop élevé pour rétablir l'asymétrie de la causalité générique. Nous nous en tiendrons donc à la thèse selon laquelle la causalité générique n'est pas, après tout, asymétrique. Les cas mentionnés par Williamson seront considérés comme des contre-exemples à la thèse de l'asymétrie.

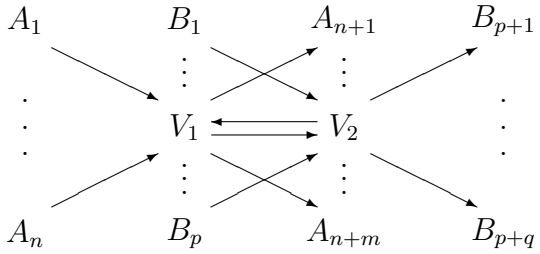
1.3.1.2 Rétablir artificiellement l'acyclicité

Nous avons montré plus haut que l'acyclicité du graphe qui représente les relations de cause à effet directes au sein d'un ensemble de variables équivaut à l'asymétrie de la relation de causalité sur cet ensemble. De ce que la causalité générique n'est pas asymétrique, il découle donc que la troisième composante de l'hypothèse de représentation est fausse. Toutefois, l'équivalence entre l'acyclicité du graphe qui représente les relations de cause à effet directes sur un ensemble et l'asymétrie de la causalité sur cet ensemble n'a été établie que sous l'hypothèse – certes restée implicite – selon laquelle une même variable ne figure qu'une fois dans un graphe bayésien donné.

Or, Williamson explique comment lever cette hypothèse permet de représenter la causalité directe au sein d'un ensemble de variables $\mathbf{V}$ au moyen d'un graphe acyclique même quand il existe dans $\mathbf{V}$ deux variables V_1 et V_2 qui se causent directement l'une l'autre. Pour faire cela, il suffit de distinguer V_1 en tant qu'il cause V_2 et V_1 en tant qu'il est causé par V_2 :

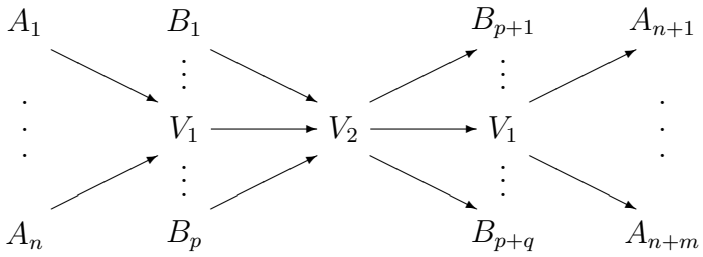
> Il est possible d'éradiquer ces cycles en considérant différentes instanciations des causes et des effets comme différentes variables : si la première augmentation des prix de l'immobilier est une variable différente de la seconde augmentation des prix, alors il existe une chaîne de connexions causales d'une augmentation à l'autre, plutôt qu'un cercle entre augmentation des prix et empressement à acheter.[44]

Plus formellement, on remplacera, dans le graphe dans lequel une variable ne figure qu'une fois et chaque relation de cause à effet directe est représentée par une flèche, chaque sous-graphe de la forme :



[44] Williamson (2004) p.50.

par un graphe de la forme :



L'artifice envisagé par Williamson permet de représenter la causalité directe sur **V** par un graphe acyclique même quand il existe deux variables de **V** qui se causent directement l'une l'autre. Cependant nous avons vu que, en l'absence d'artifice, la troisième composante de l'hypothèse de représentation implique non seulement que la causalité directe, mais encore que la causalité (tout court) est asymétrique. Pour établir que la non-asymétrie de la causalité générique n'interdit jamais de représenter la causalité directe sur un ensemble de variables au moyen d'un graphe acyclique, il nous faut donc montrer comment la causalité directe sur **V** peut être représentée par un graphe acyclique quand il existe un cycle causal impliquant au moins trois variables de **V**. Autrement dit, il convient de généraliser la proposition de Williamson en vue de rendre compte des cas où la causalité n'est pas asymétrique mais la causalité directe l'est. Il nous semble que cette généralisation est immédiate : étant donné un cycle causal, on choisit une des variables de ce cycle et on distingue cette variable en tant qu'origine du cycle de cette variable en tant qu'aboutissement du cycle. Selon les mêmes voies que celles que nous avons indiquées, on peut alors faire disparaître le cycle en le déployant.

Dans cette sous-section, nous avons montré d'une part que la causalité n'est pas, finalement, asymétrique et d'autre part que la troisième composante de l'hypothèse de représentation n'implique pas, finalement, qu'elle l'est. Si nous avons pu prouver plus haut qu'elle l'impliquait, c'est que nous avons considéré qu'une même variable ne devait figurer qu'une fois dans le graphe représentant les relations de cause à effet directes sur un ensemble de variables auquel elle appartient. En termes plus généraux, nous avons négligé ceci que la troisième composante de l'hypothèse de représentation porte sur la façon dont il est possible de *représenter* la causalité. Si la causalité était asymétrique, prendre en compte le mode de représentation le plus immédiat pour la causalité directe sur un ensemble de variables – le seul que nous ayons envisagé dans la sous-section 1.2.1 – suffirait à établir la troisième composante de l'hypothèse de représentation. Mais, ainsi que nous l'avons montré, qu'elle ne le soit pas n'implique pas que la troisième composante de

l'hypothèse de représentation est fausse. Moyennant l'artifice que nous avons présenté, il est toujours possible de représenter les relations de cause à effet directes sur un ensemble de variables au moyen d'un graphe acyclique. Nous avons traité les contre-exemples à la troisième composante de l'hypothèse de représentation considérée comme hypothèse portant nécessairement sur la causalité elle-même ; nous pouvons donc nous tourner maintenant vers la condition de Markov causale.

1.3.2 Condition de Markov causale

De même que l'hypothèse selon laquelle la causalité est asymétrique, l'hypothèse selon laquelle la causalité et les probabilités sont dans le rapport décrit par la condition de Markov causale est apparue plausible dans un premier temps. Pourtant, comme elle, elle ne peut pas être admise comme vraie après qu'on l'a soumise à un examen plus rigoureux que celui auquel nous nous sommes livrés dans la sous-section 1.2.2. De même que l'hypothèse selon laquelle la causalité est asymétrique, la condition de Markov causale admet des contre-exemples – auxquels nous venons maintenant.

1.3.2.1 Violations de la condition de Markov causale

L'essentiel des critiques qui ont été soulevées contre les réseaux bayésiens causaux porte sur la condition de Markov causale[45], et consiste souvent à lui opposer des contre-exemples. De ces contre-exemples, donc, la littérature critique sur les réseaux bayésiens causaux regorge. Dans le paragraphe qui commence, nous ne prétendons ni tracer en théorie les limites du domaine de vérité de la condition de Markov causale[46], ni même proposer une liste ou une typologie exhaustive des contre-exemples connus à la condition. Positivement, le traitement que nous donnons de ces contre-exemples est guidé par le principe suivant : donner un exemple classique pour chacun des types de violations de la condition de Markov causale que nous identifions. La raison en est double : d'une part l'argument que nous proposons dans cette sous-section prétend valoir de manière générale et indépendamment de la diversité réelle des contre-exemples à la condition de Markov causale ; d'autre part le lecteur qui souhaite prendre la mesure de cette diversité réelle peut se reporter aux textes mentionnés un peu plus haut en note.

[45] En particulier : Cartwright (1999), Cartwright (2001) section 4, Freedman et Humphreys (1999) pp. 31–34, Williamson (2001) §2, Williamson (2005) section 4.2.

[46] Nous reviendrons sur cette question et proposerons un élément de réponse dans le chapitre 4.

Pour ce qui est, maintenant, de l'organisation de la section, nous reprenons la typologie des indépendances fondamentales impliquées par la condition de Markov causale qui a servi à l'organisation de la sous-section 1.2.3. Toutefois, nous ne nous conformons pas à l'ordre de présentation adopté dans la sous-section 1.2.3. En effet, c'est d'abord et principalement la troisième sous-hypothèse de la condition de Markov causale qui a été remise en cause, les contre-exemples à la première sous-hypothèse dérivant des contre-exemples à la troisième. Dans ces conditions, le lecteur aura compris que nous envisageons les violations de la troisième sous-hypothèse de la condition de Markov causale d'abord. Nous en venons ensuite à la première composante, qui est susceptible de contre-exemples du même type que les contre-exemples à la troisième composante. La deuxième composante de l'hypothèse de représentattion est abordée dans un dernier temps.

Fourches interactives. Selon la troisième composante de la condition de Markov causale, les causes communes font écran à la dépendance entre leurs effets. Or, les exemples de causes communes, ou d'ensembles de causes communes, qui ne font pas écran entre leurs effets sont nombreux dans la littérature sur la condition de Markov causale.[47] Les premiers d'entre eux sont introduits par Salmon. Plus précisément, Salmon introduit l'idée selon laquelle il existe, à côté des fourches conjonctives reichenbachiennes, des *fourches interactives* :

> Des boules de billard reposent sur le tapis, de telle sorte que le joueur peut mettre la boule noire dans le filet à un bout de la table si et presque seulement si sa boule de choc va dans le filet à l'autre bout de table. Etant relativement novice, le joueur ne réalise pas ce fait ; par ailleurs, son habileté est telle qu'il a seulement 50 % de chances mettre la boule noire dans le filet s'il essaie. Supposons en outre que si les deux boules tombent dans leur filet respectif, la boule noire tombera avant la boule de choc. Soit A l'événement que le joueur tente le tir, B la chute de la boule noire dans le filet du bout de la table, C la chute de la boule de choc dans le filet de l'autre bout de la table. [...] L'événement A, qui doit assurément être considéré comme une cause directe à la fois de B et de C, ne fait pas écran entre B et C, puisque $P(C \mid A) = 1/2$ alors que $P(C \mid A.B) = 1$.[48]

Pour commencer, il convient de noter à la fois que Salmon parle de probabilités d'événements et que l'exemple se transpose sans mal à un cadre

[47] En particulier : Cartwright (1999) pp. 7–8, Davis (1988) p. 156, Salmon (1980) pp. 150–151, Salmon (1984) pp. 168–169.

[48] Salmon (1980) pp. 150–151.

dans lequel les probabilités sont définies sur des variables (ici des variables binaires). Maintenant, ce que Salmon fait apparaître est l'existence de situations telles qu'une cause produit un de ses effets si et seulement si elle produit l'autre. Plus précisément, Salmon écrit : « le joueur peut mettre la boule noire dans le filet à un bout de la table si et *presque* seulement si sa boule de choc va dans le filet à l'autre bout de la table »[49].

L'adverbe « presque » est important ici. Il fait apparaître qu'une cause commune A échoue à faire écran entre ses effets B et C non pas seulement si elle produit B exactement quand elle produit C, mais dès que sa production de B n'est pas indépendante de sa production de C. En termes épistémiques, la classe des contre-exemples n'est pas caractérisée par :

1. savoir que A a eu lieu ne permet pas de conclure que B ou C a eu lieu ;
2. sachant que A a eu lieu, savoir que B (resp. C) a eu lieu permet de conclure que C (resp. B) a eu lieu.

Elle est plutôt caractérisée par ceci :

1. savoir que A a eu lieu ne permet pas de conclure que B ou C a eu lieu ;
2'. sachant que A a eu lieu, apprendre que B (resp. C) a eu lieu modifie le degré de croyance en l'occurrence de C (resp. B).

Dans tous les cas ainsi caractérisés, la troisième composante de la condition de Markov causale est violée.

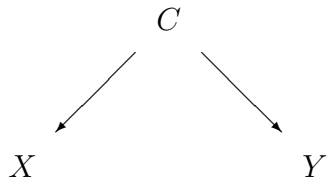
Parmi les contre-exemples à la troisième composante de l'hypothèse de Markov causale, une classe numériquement et conceptuellement importante met en scène des causes indéterministes. Nous n'abordons pas ici la question de savoir quelle est la place exacte de cette sous-classe dans la classe de tous les contre-exemples à la troisième composante de la condition de Markov causale. Cette question sera abordée plus précisément dans le chapitre 4. Nous nous contentons ici de présenter celui qui est canonique parmi les contre-exemples indéterministes à la troisième composante de la condition de Markov causale. Il est introduit par Cartwright dans les termes suivants :

> Deux usines sont en concurrence pour produire un produit chimique qui est consommé immédiatement pas une usine d'épuration proche. La ville fait une étude pour décider à laquelle faire appel. Certains jours les produits chimiques sont achetés à Clean / Green, d'autres à Cheap-but-Dirty. Cheap-but-Dirty emploie un processus véritablement probabiliste pour produire le composant. La probabilité d'obtenir le composant désiré un jour quelconque d'activité de l'usine de 0,8. Donc dans environ un cinquième des cas où le composant est

[49] Salmon (1980) p. 150. Nous ajoutons les italiques.

> acheté à Cheap-but-Dirty, les eaux usées ne sont pas traitées. Mais la méthode est si bon marché que la ville est prête à s'accommoder de cela. En revanche il existe une autre raison pour laquelle elle ne veut pas acheter à Cheap-but-Dirty : elle refuse les polluants qui sont émis en même temps que le composant dès que celui-ci est produit.[50]

Pour comprendre que ce passage introduit un contre-exemple à la troisième composante de l'hypothèse de Markov causale, considérons les variables X, Y, et C représentant respectivement la production du composant chimique par Cheap-but-Dirty, l'émission de produits polluants par Cheap-but-Dirty et le fonctionnement même de l'usine Cheap-but-Dirty. Les seules relations de cause à effet sont de C à X et de C à Y. Le graphe représentant les relations de cause à effet directes sur $\{X, Y, C\}$ est donc le suivant :



Or X et Y ne sont pas indépendantes relativement à C : $p(X = 1|Y = 1, C = 1) = 1 \neq 0,8 = p(X = 1|C = 1)$. De même que dans le cas précédent, la cause commune ne fait pas écran à ses effets parce qu'elle ne les produit pas indépendamment l'un de l'autre. L'apport spécifique de l'exemple de Cartwright est de lier les phénomènes de ce type à la notion de cause « véritablement probabiliste »[51] – ou, en d'autres termes, indéterministe. Cette notion permet de former des contre-exemples à la première sous-hypothèse de la condition de Markov causale.

Dépendance causale et indépendance probabiliste. Selon la première sous-hypothèse de la condition de Markov causale, toute variable est indépendante, relativement à l'ensemble de ses causes directes dans un ensemble de variables $\mathbf{V}$, de toutes les variables de $\mathbf{V}$ dont elle est causalement indépendante. Pour comprendre comment l'existence de causes qui ne suffisent pas à déterminer la valeur de leurs effets implique une classe de contre-exemples à cette première sous-hypothèse, il convient de montrer d'abord que deux variables causalement indépendantes peuvent être dépendantes en probabilités. L'exemple le plus fameux est ici celui qui est proposé dans Sober (1988) :

[50] Cartwright (1999) p. 7.
[51] Cartwright (1999) p. 7.

> Considérons le fait que le niveau de la mer à Venise et le coût du pain en Grande-Bretagne ont été tous deux à la hausse dans les deux siècles passés. Disons que tous deux ont augmenté de façon monotone. Imaginons que nous mettions ces informations sous la forme d'une liste chronologique. Pour chaque date, nous relevons le niveau de la mer à Venise et le prix courant du pain britannique. Parce que les deux quantités ont augmenté régulièrement avec le temps, il est vrai que les niveaux de la mer plus élevés que la moyenne tendent à être associés à des prix du pain plus élevés que la moyenne. Les deux quantités sont très fortement corrélées positivement.
>
> Il me semble que nous ne nous sentons pas conduits à expliquer cette corrélation par une cause commune. Plutôt, nous condidérons les niveaux de la mer à Venise et les prix du pain britanniques comme augmentant tous deux pour des raisons endogènes et quelque peu isolées. Les conditions locales vénitiennes ont augmenté le niveau de la mer et des conditions locales assez différents ont poussé à la hausse le coût du pain en Grande-Bretagne. Ici, postuler une cause commune est simplement peu plausible, étant donné le reste de ce que nous croyons.[52]

Ainsi qu'il est explicite dans cet extrait, l'exemple de Sober est conçu comme un contre-exemple au principe de la cause commune. Néanmoins, pas plus qu'une cause commune, nous ne sommes prêts à reconnaître un quelconque autre lien causal entre le niveau de la mer à Venise et le prix du pain britannique. Nous considérons donc le cas envisagé comme une illustration de la possibilité qu'existe une dépendance probabiliste sans dépendance causale de quelque forme qu'elle soit.

Imaginons maintenant que le prix du pain en Grande-Bretagne soit complètement déterminé par les valeurs des variables d'un ensemble $\mathbf{V}$ – auquel appartiendraient en particulier les variables représentant le prix du blé et la demande de pain en Grande-Bretagne. Dans ce cas, le prix du pain britannique est indépendant du niveau de la mer à Venise relativement à l'ensemble de ces variables. Maintenant, on a vu dans le paragraphe précédent qu'il existe des causes dont la valeur ne suffit pas à déterminer celle de leur effet. Si c'est le cas de l'ensemble des causes du prix du pain britannique, alors celui reste dépendant du niveau de la mer à Venise, même quand on conditionnalise sur l'ensemble de ses causes. En s'appuyant à la fois sur l'existence de dépendances probabilistes auxquelles ne correspondent pas de dépendances causales et sur l'existence de causes probabilistes, on fait donc apparaître une classe de contre-exemples à la première sous-hypothèse de la condition de Markov causale.

[52] Sober (1988) p. 215.

Processus non-markoviens Selon la deuxième composante de la condition de Markov causale, une variable est indépendante de ses causes indirectes relativement à ses causes directes. En d'autres termes, ses causes immédiates d'une variable font écran entre cette variable et ses ancêtres causaux. Cette composante proprement markovienne de la condition de Markov causale semble avoir pour contre-exemples tous les processus non markoviens. « Processus non markoviens » s'entend ici au sens historiquement premier de l'expression. En ce sens, un processus non markovien affectant un système donné est tel que l'état du système à un instant donné ne suffit pas à déterminer les probabilités que le système soit dans tel ou tel état à l'instant immédiatement postérieur. Les probabilités à l'instant t_{n+1} ne dépendent pas seulement de ce qui à eu lieu à l'instant t_n ; elles dépendent aussi de ce qui a eu lieu avant. Les phénomènes d'apprentisssage, qu'ils soient individuels ou collectifs, sont couramment non-markoviens en ce sens.[53] Si les phénomènes non-markoviens en ce sens sont bien des contre-exemples à la condition de Markov causale, alors il s'ensuit une limitation sensible du domaine au sein duquel les réseaux bayésiens causaux sont un outil d'analyse pertinent.

En vue de comprendre en quel sens les processus non markoviens au sens classique que nous venons de rappeler constituent des contre-exemples à la condition de Markov causale, il convient de prendre un exemple. Celui que nous choisissons est extrêmement simple et épuré ; il fonde d'ailleurs une classe importante de modèles pour les processus non-markoviens. L'exemple est le suivant : celui d'une urne qui contient à l'instant initial t_0 une boule blanche et une boule noire, et dans laquelle on effectue une suite de tirages avec double remise : à la suite du tirage, on remet dans l'urne *deux* boules de la même couleur que la boule tirée. La suite des résultats de tirages effectués dans l'urne est un processus non-markovien : la probabilité de tirer une boule blanche (resp. noire) à l'instant t_{n+1} dépend non seulement du résultat du tirage effectué à l'instant t_n, mais encore des résultats de tous les tirages antérieurs.

La variable pour laquelle les probabilités ne sont pas complètement déterminées par la valeur prise par ses parents immédiats est la variable binaire T_{n+1} qui représente le résultat du tirage effectué à l'instant t_{n+1}. Cette variable a un seul parent immédiat : la variable binaire T_n qui représente le résultat du tirage effectué à l'instant t_n. De façon plus générale, le processus non-markovien constitue un contre-exemple à la condition de Markov qui définit les réseaux bayésiens quand on le représente de la façon suivante :

$$T_1 \longrightarrow T_2 \longrightarrow \ldots \longrightarrow T_n \longrightarrow T_{n+1} \longrightarrow \ldots$$

[53]Pour une analyse des phénomènes sociaux de renforcement, voir Skyrms (2004).

où T_i est la variable binaire qui prend la valeur B ou la valeur N selon que la boule tirée à l'instant t_i est blanche ou noire. Le couple composé du graphe que nous venons de tracer et des probabilités classiques[54] n'est pas un réseau bayésien.

Il apparaît alors que le processus non-markovien que nous envisageons constitue un contre-exemple à la condition de Markov causale si et seulement si on considère que les flèches du graphe que nous avons tracé sont causales. En d'autres termes, la suite des tirages avec double remise invalide la condition de Markov causale si et seulement si chaque variable T_{n+1} a pour cause directe dans $\mathbf{T}$ la seule variable T_n. Or il nous semble que nous ne serions pas prêts à dire cela. En effet, si on reconnaît que T_n est une cause directe de T_{n+1}, c'est parce que la valeur que prend T_n affecte la composition de l'urne, dont dépendent les probabilités sur T_{n+1}. Mais, alors, il n'y a aucune raison de ne pas considérer les variables T_1 à T_{n-1} aussi comme des causes directes de T_n. Or, clairement, cela suffit à rétablir la condition de Markov. L'exemple que nous envisageons ne donne donc pas de contre-exemple à la condition de Markov *causale*. Positivement, il constitue un contre-exemple à ce que serait la condition de Markov sous une interprétation *temporelle* des réseaux bayésiens. En effet, les flèches du graphe que nous avons construit représentent les relations d'antériorité temporelle immédiate : une flèche de T_i à T_{i+1} signifie que T_i précède immédiatement T_{i+1} dans le temps.

De façon plus générale, les processus non-markoviens au sens classique ne constituent des contre-exemples à la condition de Markov causale qu'à la condition de considérer que la seule cause de ce qui advient à un instant donné est ce qui advient à l'instant immédiatement précédent. Or, justement, ces processus non-markoviens sont tels qu'il existe une action du passé vers le futur indépendante de l'action du présent sur le futur. Il nous semble même que c'est ce qui les caractérise : ce qui fait que les processus d'apprentissage sont non-markoviens est précisément que le passé a un effet sur le futur, indépendamment du présent. Dans ces conditions, les processus non-markoviens au sens classique ne sont pas des contre-exemples à la deuxième composante de la condition de Markov causale. En outre, nous ne connaissons pas d'autres candidats au titre de contre-exemples à cette deuxième composante. En conséquence, nous nous en tiendrons aux contre-exemples aux première et troisième composantes de la condition de Markov causale que nous avons discutés plus haut.

Les contre-exemples à la première et, surtout, à la troisième composante

[54]Nous entendons par là que la probabilité de B (resp. N) à l'occasion du n+1-ème tirage est le nombre de boules blanches dans l'urne à l'issue du n-ième tirage divisé par le nombre total de boules dans l'urne à l'issue du n-ième tirage.

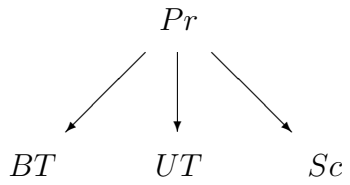
impliquent que la condition de Markov causale ne vaut pas dans le cas général. Par ailleurs, ces contre-exemples ne semblent pas pouvoir être intégrés au domaine de pertinence des réseaux bayésiens causaux de la même façon que l'ont été les cas de cycles causaux : si la causalité et les probabilités sur un ensemble de variables ne sont pas toujours dans un rapport markovien, on ne peut rien y changer. C'est cette analyse que nous critiquons dans le prochain paragraphe. Nous y mettons en évidence un artifice analogue de celui qui a été discuté plus haut.

1.3.2.2 Rétablir artificiellement la condition de Markov causale

Avant de montrer comment la condition de Markov causale peut être rétablie artificiellement selon des voies analogues de celles que Williamson préconise pour l'acyclicité, il convient d'abord de pointer un moyen naturel de prendre en charge certains contre-exemples à la condition de Markov causale. Ces contre-exemples sont du type de celui-ci, introduit par Gillies :

> Le second contre-exemple pourrait être appelé exemple de la vache enceinte et est donné dans Jensen (1996, pp.36–37). Il vient du domaine de la science vétérinaire, et est relatif à un test destiné à déterminer si une vache est enceinte. [...]
> Ici Pr est une variable qui prend la valeur 1 si la vache est enceint et la valeur 0 sinon. BT représente le résutat d'un test sanguin, UT le résultat d'un test d'urine, et Sc le résultat d'un scanner. [...] Les conditions d'indépendance [imposées par la structure causale] sont ici les suivantes : les variables BT, UT et Sc doivent être indépendantes relativement à Pr. Maintenant Sc est en effet indépendante de BT et de UT relativement à Pr, mais BT et UT sont corrélées relativement à Pr.[55]

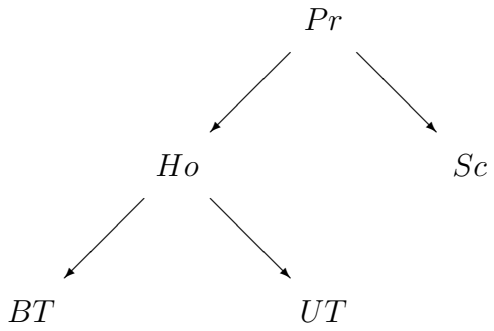
La structure causale sur l'ensemble de variables envisagé par Gillies est la suivante :



Maintenant, le problème est que UT et BT ne sont pas indépendantes relativement à Pr comme l'exige la condition de Markov causale. Toutefois, continue Gillies, chacune de ces deux variables est un effet de la variable Ho

[55] Gillies (2002) p. 80.

représentant le niveau hormonal de la vache et cette variable fait écran entre UT et BT. On rétablit donc la condition de Markov causale en introduisant Ho à l'ensemble de variables considéré. Ainsi, le graphe



est un graphe bayésien.

La violation de la condition de Markov décrite par Gillies et la façon dont il envisage de rétablir la condition dépendent de ceci qu'un réseau bayésien est défini sur un ensemble de variables fini donné. Dans l'exemple de la vache enceinte, la condition de Markov causale est violée parce que la variable Ho n'a pas été prise en compte, et on rétablit la condition en intégrant Ho à l'ensemble de variables considéré. Maintenant, les contre-exemples à la condition de Markov causale que nous avons présentés dans le paragraphe précédent ne sont pas de ce type. Ils ne dépendent pas de l'omission d'une variable, mais du caractère « véritablement probabiliste » – pour reprendre les mots de Cartwright – de l'action de certaines causes sur leurs effets. Il apparaît alors que les contre-exemples que nous avons envisagés continuent d'exister quand toutes les variables pertinentes pour la structure causale de la situation considérée sont prises en compte.

Les variables pertinentes auxquelles nous faisons allusion ici sont des variables observables. Mais on peut imaginer revenir sur cette caractéristique. Pour le dire autrement, on peut imaginer rétablir la condition de Markov causale dans les cas qui nous intéressent en introduisant une variable *non observable* qui joue exactement le même rôle que joue Ho dans l'exemple de la vache enceinte. Ainsi, dans le cas de l'usine Cheap-but-Dirty, on introduirait une variable E représenter l'efficacité du fonctionnement de l'usine. De la même façon que Ho fait écran entre BT et UT dans l'exemple de la vache enceinte, cette variable fait écran entre X et Y dans l'exemple de Cheap-but-Dirty.

La stratégie s'étend au contre-exemple que nous avons construit pour la première sous-hypothèse de la condition de Markov causale. Il semble en effet suffisant d'introduire une variable représentant le temps pour rendre

indépendants le prix du pain en Grande-Bretagne et le niveau de la mer à Venise. Toutefois, on voit mal comment faire dans le cas des boules de billard et, surtout, rien ne garantit que la stratégie puisse être généralisée. Pour surmonter cette difficulté, il convient de dissocier entre identification et définition d'une variable. Ce qui est vrai des deux cas que nous avons présentés, c'est qu'ils sont tels que la variable à introduire s'identifie facilement parce qu'elle se définit aisément dans le langage naturel – à défaut d'être observable. En revanche il est faux que la propriété d'être aisément définissable dans le langage naturel soit co-extensive de la propriété d'être définissable tout court. Ainsi, même dans les cas où nous ne savons pas dire correctement quelle variable doit être introduite pour que la condition de Markov causale ne soit plus violée, il est possible d'introduire une telle variable. De façon générique, elle sera décrite comme un « noeud caché ». Kwoh et Gillies proposent une méthode pour définir sur l'ensemble de variables étendu une distribution de probabilités qui 1) conserve les distributions de probabilités marginales sur les sous-ensembles de l'ensemble de variables initial ; 2) étend la distribution de probabilités conditionnelle initiale ; 3) est telle que la condition de Markov est satisfaite pour le graphe et la distribution de probabilités étendus.[56]

Il apparaît donc que les contre-exemples à la condition de Markov causale peuvent tous être intégrés au domaine de pertinence des réseaux bayésiens causaux quand on remet en cause l'hypothèse naturelle selon laquelle le graphe représentant les relations de cause à effet directes sur un ensemble de variables observables a pour sommets seulement des variables de ce type. De façon exactement similaire, c'est en remettant en cause l'hypothèse selon laquelle une même variable ne peut figurer qu'une fois dans un graphe bayésien que Williamson résout le problème posé par les cycles causaux.

Reste, maintenant, que l'introduction de variables a souvent été proposée comme solution à des violations de la condition de Markov causale[57] et que, à ce type de solutions, on oppose généralement que les variables ainsi introduites ne représentent rien qui soit susceptible d'entrer dans une relation de cause à effet.[58] Formulée autrement, la critique consiste à dire que le graphe sur l'ensemble de variables étendu n'est plus causal. Il nous semble que cette critique n'est pas rhédibitoire si l'on revient à l'idée selon laquelle les réseaux bayésiens causaux sont un mode de représentation de la causalité directe sur un ensemble de variables observables. Etant donné un ensemble de variables observables $\mathbf{V}$, on peut toujours représenter les relations de cause à effet directes sur $\mathbf{V}$ au moyen d'un graphe bayésien sur $\mathbf{V}$ et si besoin un ensemble

[56] Kwoh et Gillies (1996) section 3.

[57] Spirtes, Glymour et Scheines (1993) pp.32–37 ; Pearl (2000) p. 62 .

[58] En particulier : Cartwright (2001) p. 259 ; Williamson (2001) §2.

V' de variables non observables de la façon suivante :
- une fléche d'une variable de **V** à une variable de **V** représente une relation de cause à effet directe entre ces deux variables ;
- une flèche d'une variable $V\prime$ de **V'** à une variable V_1 de **V**
  - représente une relation de cause à effet directe entre V_2 et V_1 s'il existe V_2 de **V** qui est un parent de V' ;
  - ne représente aucune relation de cause à effet s'il n'existe pas de tel V_2.

Dans le cas de la condition de Markov causale comme dans celui de la troisième composante de l'hypothèse de représentation, renoncer au mode le plus immédiat de représentation de la causalité directe par un graphe orienté acyclique permet de sauver l'hypothèse corrélative de la notion de réseau bayésien causal. Ce n'est pas parce que les hypothèses ne valent pas toujours pour ce mode naturel de représentation que la causalité directe sur un ensemble de variables observables ne peut pas être toujours représentée au moyen d'un graphe bayésien. Nous avons décrit dans le début de cette section les voies selon lesquelles elle peut toujours l'être.

Spohn (2001) donne une justification originale de l'universalité de cette possibilité. Selon ce texte, en effet, « il n'y a rien de plus que les réseaux bayésiens dans la dépendance causale »[59] :

> c'est la structure de réseaux bayésiens convenablement raffinés qui décide ce qu'il en est des dépendances causales. Nous ne pouvons pas considérer B comme dépendant causalement de A à moins que nous trouvions une suite de flèches ou d'arêtes dirigées allant de A à B dans un réseau bayésien convenablement raffiné et à moins que, bien sûr, cela persiste pour des raffinements plus grands.[60]

Mais ce qui nous intéresse ici est moins de savoir *pourquoi* on peut toujours rétablir l'acyclicité et la condition de Markov causale que si on peut toujours le faire *effectivement*. La question est donc celle du domaine au sein duquel les artifices que nous avons décrits sont effectivement utilisables.

1.3.3 Domaine d'utilisation effective des artifices

Pour dessiner les contours du domaine dans lequel les artifices que nous avons définis peuvent être utilisés, il convient de revenir aux deux types d'utilisations des réseaux bayésiens causaux que nous avons décrits dans la sous-section 1.2.1 : d'une part l'utilisation du graphe causal afin de construire

[59] Il s'agit de la traduction du titre de Spohn (2001).

[60] Spohn (2001) p. 10. Les italiques sont dans le texte original.

un graphe qui représente la distribution de probabilités sur un ensemble de variables, d'autre part l'interprétation causale des résultats d'algorithmes de construction de graphes qui représentent des distributions de probabilités. Une analyse rapide fait apparaître que les violations de l'hypothèse d'acyclicité et de la condition de Markov causale n'ont pas le même statut dans l'un et l'autre contexte.

Dans le premier contexte, les graphes causaux sont utilisés comme des guides pour la construction de graphes bayésiens. Pour reprendre les termes de Gillies, dans ce premier contexte les graphes causaux sont des « guide heuristique pour la construction de réseaux bayésiens »[61]. Si l'hypothèse d'acyclicité ou la condition de Markov sont violées, ce guide est trompeur : les graphes causaux ne sont pas des graphes bayésiens. Il convient alors de ne pas suivre ce guide, ou plutôt de ne pas le suivre aveuglément. On peut toutefois prendre ce guide – le graphe causal – pour point de départ. Plus précisément, on procédera de la façon suivante : 1. construire le graphe causal, 2. déterminer s'il est bayésien et 3. l'amender en vue de le rendre bayésien s'il ne l'est pas déjà.[62] Les artifices que nous avons présentés ont toute leur place en 3.

Pour ce qui est, maintenant, du second contexte dans lequel les réseaux bayésiens apparaissent, les choses se présentent différemment. Rappelons en effet que le graphe causal n'est pas connu initialement, mais qu'il est ce qu'on prétend inférer, sous l'hypothèse qu'il représente la distribution de probabilités sur l'ensemble de variables considéré. On suppose donc que le graphe causal est un graphe bayésien. Or un graphe bayésien sur un ensemble de variables $\mathbf{V}$ est tel qu'une même variable n'y figure qu'une fois. Il en découle que l'artifice proposé par Williamson pour les cas dans lesquels la causalité directe n'est pas asymétrique ne peut pas être utilisé. Si la causalité directe n'est pas *effectivement* asymétrique sur l'ensemble de variables $\mathbf{V}$, les algorithmes considérés ne produisent pas des représentations correctes de la causalité directe sur $\mathbf{V}$.

Le cas de la condition de Markov causale semble plus complexe que celui de la troisième composante de l'hypothèse de représentation. En effet, les plus perfectionnés des algorithmes que nous discutons ici (IC*, CI, FCI) prennent en compte la possibilité qu'existent des causes communes à des variables de $\mathbf{V}$ qui n'appartiennent pas elles-mêmes à $\mathbf{V}$. La question qui se pose est celle du rapport que cette prise en compte entretient avec le second des artifices que nous avons définis plus haut. Or, une analyse rapide fait apparaître que

[61] Gillies (2002) p. 78.

[62] Pour une présentation plus exhaustive de la méthode qu'il appelle « méthode qual-quant », voir Gillies (2002) pp. 84–85.

les deux procédures ne visent pas les mêmes cas. L'artifice que nous avons défini visait à intégrer au domaine de ce qui est représentable par les réseaux bayésiens causaux, les cas tels que soit il n'existe pas de relation causale entre deux variables pourtant dépendantes, soit deux variables dont aucune des deux n'est cause de l'autre ne sont pas indépendantes relativement à une cause qui leur est commune. De l'autre côté, les perfectionnements des algorithmes auxquels nous faisions allusion ne visent qu'à prendre en compte le fait que certaines causes communes à des variables de l'ensemble de variables observées $\mathbf{V}$ peuvent ne pas appartenir elles-mêmes à $\mathbf{V}$ – exactement de la même façon que la variable Ho avait été initialement omise dans l'exemple de la vache enceinte. Ce n'est donc pas la possibilité de violations de la condition de Markov causale, au sens fort que nous avons donné à ce terme, qui est prise en compte par là. Il n'y a donc pas de sens, dans ce type d'utilisations des réseaux bayésiens causaux, à recourir à l'artifice que nous avons défini plus haut. De façon plus générale, les algorithmes d'inférence aux causes qui mobilisent les réseaux bayésiens donnent des résultats corrects seulement quand l'hypothèse de représentation et la condition de Markov causale sont effectivement satisfaites pour la représentation graphique naturelle des relations de cause à effet directes sur $\mathbf{V}$. Négativement, les contre-exemples du type de ceux que nous avons présentés dans la section précédente sont des cas dans lesquels l'inférence aux causes fondée sur la notion de réseau bayésien causal n'est pas valide.

1.4 Conclusion

Dans la première section de ce chapitre, nous avons expliqué dans quel contexte théorique les réseaux bayésiens apparaissent, nous les avons définis, et nous avons présenté leurs principales propriétés formelles. Dans la suite, nous avons concentré notre attention sur l'interprétation causale des réseaux bayésiens, et plus précisément sur les hypothèses corrélatives de la notion de réseau bayésien causal.

La deuxième section a fait apparaître d'abord que, seules parmi ces hypothèses, l'hypothèse d'acyclicité et la condition de Markov causale ont des implications relativement à la causalité elle-même, et ensuite que ces hypothèses se présentent comme plausibles.

Ces deux résultats ont été remis en cause dans la troisième et dernière section. D'abord nous avons présenté des contre-exemples à la troisième composante de l'hypothèse de représentation et à la condition de Markov causale. Ensuite, nous avons montré que ces contre-exemples n'en sont qu'à la condition de s'en tenir au mode naturel de représentation de la causalité directe par

un graphe acyclique orienté. Autrement dit, nous avons montré que les deux hypothèses en question n'ont d'implications relatives à la causalité elle-même que si l'on adopte ce mode de représentation. Positivement, nous avons défini des artifices permettant de toujours représenter la causalité directe sur un ensemble de variables au moyen d'un graphe bayésien. Finalement, nous avons montré que ces artifices ne peuvent pas toujours être effectivement utilisés. Plus précisément, nous avons fait apparaître que le statut des hypothèses et, avec lui, la réponse à la question de la possibilité d'utiliser effectivement les artifices définis varient selon le contexte théorique dans lequel on utilise les réseaux bayésiens causaux. Ils peuvent être utilisés quand le graphe causal, connu, est un guide dans la construction d'un graphe bayésien ; ils ne peuvent pas l'être quand le graphe causal est précisément ce que vise l'inférence.

Ainsi qu'il doit être clair à ce point de notre travail, c'est le second type d'utilisations qui nous intéresse dans cette première partie. Pour ce qui concerne ces utilisations, les résultats établis dans le chapitre qui s'achève nous permettent d'en venir aux questions qui nous intéressent. Ainsi, ils nous permettent d'en venir à la question des modalités de l'inférence aux causes quand elle est fondée sur les réseaux bayésiens causaux. Nous avons vu que cette question se pose d'abord du point de vue de l'analyse conceptuelle, ou plus précisément du point de vue du rapport entre l'épistémologie et l'analyse conceptuelle. La question est alors celle du rapport entre le critère de causalité que véhiculent les réseaux bayésiens causaux et les théories probabilistes de la causalité. Elle est traitée dans le prochain chapitre.

Appendice
Définitions en théorie des graphes et en théorie des probabilités

Notions de théorie des graphes

Graphes

Définition 1.11 (Graphe) *Un* graphe G *est un couple* $(\boldsymbol{V}, \boldsymbol{L})$ *où* $\boldsymbol{V}$ *est un ensemble de variables et* $\boldsymbol{L}$ *un ensemble de paires (ordonnées ou non) d'éléments de* $\boldsymbol{V}$.

Terminologie. Soit $G = (\mathbf{V}, \mathbf{L})$.
On dit que :
- les éléments de $\mathbf{V}$ sont les *sommets* ou *noeuds* de G ;
- les éléments de $\mathbf{L}$ sont les *liens* de $\mathbf{V}$;
- G est un graphe *sur* $\mathbf{V}$;
- deux éléments A et B de $\mathbf{V}$ sont *adjacents dans* G si la paire (A, B) ou la paire (B, A) – qu'elles soient ou non ordonnées – appartiennent à $\mathbf{L}$;
- une suite d'éléments de V est un *chemin* de G si chaque variable qui apparaît dans la suite est adjacente dans G à son successeur dans la suite s'il existe.

Graphes orientés

Liens orientés. Un lien est *orienté* s'il est une paire ordonnée ; il est alors représenté par une flèche. Dans le cas contraire, il est non orienté et représenté par un simple trait, aussi appelé *arête*.

Graphes orientés. Un graphe dont tous les liens sont orientés est lui-même orienté ; un chemin tel que chaque variable A qui y figure constitue avec son successeur B, s'il existe, un lien orienté (A, B) est lui-même orienté.[63]

Graphes orientés cycliques et acycliques. Un chemin orienté est un *cycle* quand le premier élément du premier lien et le second élément du dernier lien qui le constituent sont identiques. Un graphe orienté est *acyclique* quand il ne comporte pas de cycle.

[63] Remarquons que, sous ces définitions, il existe des chemins non orientés de graphes orientés. Un exemple simple est le chemin (A, B, C) du graphe orienté $A \longrightarrow B \longleftarrow C$.

Terminologie. On utilise généralement la terminologie de la parenté pour désigner les relations entre les variables d'un graphe acyclique orienté. En particulier :

- l'ensemble des *parents* d'une variable A de V, noté $\mathbf{PA}(A)$, est l'ensemble des variables de V dont part une flèche qui pointe vers A. Un parent d'une variable a cette variable pour *enfant* ;
- l'ensemble des *ancêtres* d'une variable A de V est l'ensemble des variables de V dont part un chemin orienté qui pointe vers A. Un ancêtre d'une variable a cette variable pour *descendant*.

Notions de théorie des probabilités

Dans toute cette section, $\mathbf{V} = \{V_1, V_2, \ldots V_n\}$ est un ensemble de variables aléatoires discrètes susceptibles de prendre chacune un nombre fini de valeurs.

Distribution de probabilités

Valeur d'un ensemble de variables. Une *valeur de* $\boldsymbol{V}$ est une conjonction de la forme « V_1 prend la valeur v_1 et V_2 prend la valeur v_2 et … et V_n prend la valeur v_n ». On note $(v_1, v_2, \ldots, v_n)$ une telle conjonction[64] et $Val(\mathbf{V})$ l'ensemble des valeurs de $\mathbf{V}$.

Définition 1.12 (Distribution de probabilités) *Une distribution de probabilités p sur $\boldsymbol{V}$ est une fonction de l'ensemble des valeurs de $\boldsymbol{V}$ dans l'intervalle réel [0 ; 1] telle que $\sum_{\boldsymbol{v} \in Val(\boldsymbol{V})} p(\boldsymbol{v}) = 1$.*

Distribution de probabilités marginale

Compatibilité. Une valeur $\mathbf{v}$ de $\mathbf{V}$ est *compatible* avec une valeur $\mathbf{w}$ d'un sous-ensemble $\mathbf{W}$ de$\mathbf{V}$ si $\mathbf{v}$ et $\mathbf{w}$ coïncident pour toutes les variables de $\mathbf{W}$. On note $Comp_{\mathbf{V}}(\mathbf{w})$ l'ensemble des valeurs de $\mathbf{V}$ compatibles avec $\mathbf{w}$.

Définition 1.13 (Distribution de probabilités marginale) *Pour tout sous-ensemble $\boldsymbol{W}$ de $\boldsymbol{V}$, la distribution de probabilités marginale sur $\boldsymbol{W}$ est la fonction $q_{\boldsymbol{W}}$ de l'ensemble des valeurs de $\boldsymbol{W}$ dans l'intervalle réel [0 ; 1] telle que : pour toute valeur $\boldsymbol{w}$ de $\boldsymbol{W}$, $q(\boldsymbol{w}) = \sum_{\boldsymbol{v} \in [Val(\boldsymbol{V}) \cap Comp_{\boldsymbol{V}}(\boldsymbol{w})]} p(\boldsymbol{v})$.*

[64] Cette notation, quoique usuelle, est trompeuse : en reprenant la notation traditionnelle pour la notion de suite, elle laisse penser que l'ordre des v_i importe – ce qui n'est évidemment pas le cas puisque les propositions qui font référence aux v_i sont conjointes.

Extension de p. La fonction qui à toute valeur $\mathbf{w}$ d'un sous-ensemble $\mathbf{W}$ de $\mathbf{V}$ associe $q(\mathbf{w})$ est une extension de p.
Pour cette raison traditionnellement notée également « p ».
De sa définition, il découle en particulier :

Proposition 1.5 *Pour toute distribution de probabilités p sur un ensemble de variables $\boldsymbol{V}$ et toute variable V de $\boldsymbol{V}$, $\sum_{v \in Val(V)} p(v) = 1$.*

Distribution de probabilités conditionnelles

Définition 1.14 (Distribution de probabilités conditionnelles) *Une distribution de probabilités conditionnelles induite par la distribution de probabilités p sur $\boldsymbol{V}$ est une fonction r de l'ensemble des paires de valeurs de sous-ensembles de $\boldsymbol{V}$ dans l'intervalle réel [0 ; 1] qui à tout couple $(\boldsymbol{t}, \boldsymbol{u})$ de $\boldsymbol{T} \times \boldsymbol{U} \subseteq \boldsymbol{V} \times \boldsymbol{V}$ associe $r(\boldsymbol{t}|\boldsymbol{u})$ tel que :*

1. $r(\boldsymbol{t}|\boldsymbol{u}).p(\boldsymbol{u}) = p(\boldsymbol{t}, \boldsymbol{u})$;
2. $\sum_{\boldsymbol{t} \in Val(\boldsymbol{T})} r(\boldsymbol{t}|\boldsymbol{u}) = 1$.

Univocité. $r(\mathbf{t}|\mathbf{u})$ est déterminée univoquement par p si et seulement si $p(\mathbf{u}) \neq 0$.

Extension de p. r peut être considérée comme une extension de p, et est donc elle aussi traditionnellement notée « p ».
De sa définition, il découle :

Proposition 1.6 *Pour toute distribution de probabilités p sur un ensemble de variables $\boldsymbol{V}$, toute distribution de probabilités conditionnelles induite par p, toute variable V de $\boldsymbol{V}$, tout sous-ensemble $\boldsymbol{W}$ de $\boldsymbol{V}$ et toute valeur $\boldsymbol{w}$ de $\boldsymbol{W}$, $\sum_{v \in Val(V)} p(v|\boldsymbol{w}) = 1$.*

Indépendances probabilistes

Définition 1.15 (Indépendance probabiliste relative) *Etant donnés trois sous-ensembles $\boldsymbol{X}$, $\boldsymbol{Y}$ et $\boldsymbol{Z}$ d'un ensemble $\boldsymbol{V}$ de variables et une distribution de probabilités sur $\boldsymbol{V}$,*
$\boldsymbol{X}$ et $\boldsymbol{Y}$ sont indépendants relativement à *$\boldsymbol{Z}$ pour p si pour tout triplet $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ de valeurs de $\boldsymbol{X}$, $\boldsymbol{Y}$ et $\boldsymbol{Z}$ respectivement, on a $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{x}|\boldsymbol{z})$.*

Convention. Dans le cas où deux ensembles de variables $\mathbf{X}$ et $\mathbf{Y}$ sont indépendants relativement à un ensemble de variables $\mathbf{Z}$ qui est un singleton $\{Z\}$, on pourra dire que $\mathbf{X}$ et $\mathbf{Y}$ sont indépendants relativement à $\{Z\}$.

Propriété. La relation d'indépendance probabiliste est symétrique :
Si $\mathbf{X} \amalg \mathbf{Y}|\mathbf{Z}$, alors $\mathbf{Y} \sqcup \mathbf{X}|\mathbf{Z}$.

Parents markoviens

Dans cette sous-section, V_i est une variable de $\mathbf{V}$, p une distribution de probabilités sur $\mathbf{V}$ et $<$ un ordre strict sur $\mathbf{V}$.

Définition 1.16 (Parents markoviens) *Un ensemble $\boldsymbol{PM}_i$ de* parents markoviens *de V_i pour p et $<$ dans $\boldsymbol{V}$ est sous-ensemble de $\boldsymbol{V}$ minimal parmi ceux qui ont les propriétés suivantes :*

- *tous les éléments de $\boldsymbol{PM}_i$ sont des prédécesseurs de V_i pour $<$;*
- *V_i est indépendant pour p de $\boldsymbol{V} \setminus \boldsymbol{PM}_i$ relativement à $\boldsymbol{PM}_i$.*

Reformulation. Un ensemble $\mathbf{PM}_i$ de parents markoviens de V_i pour p et $<$ est un sous-ensemble de $\{V_1, \ldots, V_{i-1}\}$ tel que :

1. pour toute valeur $\mathbf{pm}_i$ de $\mathbf{PM}_i$ et toute valeur $(v_1, \ldots, v_{i-1})$ de $(V_i, \ldots, V_{i-1})$ compatible avec $\mathbf{pm}_i$, $p(v_i|\mathbf{pm}_i) = p(v_i|v_1, \ldots, v_{i-1})$;
2. aucun sous-ensemble propre de $\mathbf{PM}_i$ ne satisfait la condition précédente.

Chapitre 2

Réseaux bayésiens causaux et théories probabilistes de la causalité

Maintenant que nous avons présenté les réseaux bayésiens causaux, nous pouvons en venir aux méthodes d'inférence aux causes génériques qu'ils fondent. Dans le chapitre qui commence, la question est abordée du point de l'analyse conceptuelle. Elle est alors celle de la caractérisation de la causalité qui sous-tend la définition des réseaux bayésiens causaux. Plus précisément, il s'agit pour nous ici de comparer le critère de causalité qui est à l'oeuvre dans les méthodes d'inférence causale fondées sur les réseaux bayésiens, avec les théories probabilistes de la causalité relativement auxquelles notre travail se comprend.

Cette comparaison prend place dans le projet général d'exploration des corrélats épistémologiques des théories probabilistes de la causalité. Elle se justifie par ailleurs de manière spécifique. En effet, ainsi que nous l'avons indiqué déjà et que nous le montrerons bientôt, nos meilleures théories probabilistes de la causalité sont circulaires – au sens où la notion de cause apparaît dans l'*analysans* de l'expression « A cause B ». On se demande bien, alors, comment les réseaux bayésiens permettent d'inférer des causes à partir de données de nature probabiliste.

Le chapitre qui commence se déroule en trois temps. Dans la première section, nous mettons au jour et analysons la caractérisation de la causalité qui est à l'oeuvre dans l'inférence aux causes fondée sur les réseaux bayésiens. Cette caractérisation est désignée dans la suite au moyen de l'expression « caractérisation RB ». Dans la deuxième section, nous présentons les théories probabilistes de la causalité. Enfin, la troisième section est consacrée à la comparaison elle-même des caractérisations probabilistes de la causalité.

2.1 Caractérisation RB de la causalité

Ainsi que nous venons de l'annoncer, la section qui commence ici porte sur la caractérisation de la causalité qui est à l'oeuvre dans les méthodes d'inférence aux causes fondées sur les réseaux bayésiens. Ce qui doit être entendu par là demande à être précisé. En vue de cette précision, rappelons que nous visons la comparaison de cette caractérisation avec les théories probabilistes de la causalité. Ce que nous entendrons par « caractérisation de la causalité » doit donc pouvoir être comparé avec les théories probabilistes de la causalité. Autrement dit, il doit s'agir d'une condition nécessaire et suffisante pour que « X cause Y » soit vraie dans le cadre théorique constitué par les réseaux bayésiens causaux.

La section s'organise de la façon suivante : après avoir présenté la méthode que nous adoptons pour ceci, nous mettons au jour la caractérisation RB de la causalité. Une de ses conséquences est explorée dans la troisième sous-section.

2.1.1 Méthodologie

Deux voies semblent s'offrir à nous afin de mettre au jour les conditions pour la causalité qui sont à l'oeuvre dans le contexte de l'inférence aux causes génériques fondée sur les réseaux bayésiens. D'un côté, on peut revenir à la définition même des réseaux bayésiens causaux et s'intéresser aux corrélats probabilistes d'une flèche dans un graphe bayésien sur un ensemble de variables qui satisfait les hypothèses requises par l'inférence aux causes fondée sur les réseaux bayésiens. D'un autre côté, on peut considérer directement les algorithmes d'inférence aux causes fondés sur les réseaux bayésiens (« algorithmes RB » dans la suite) que nous avons mentionnés dans la sous-section 1.2.1. On s'attacherait alors à déterminer à quelle(s) condition(s) ils ont pour résultat un graphe dans lequel figure une flèche d'une variable X à une variable Y de l'ensemble de variables considéré.

En vue de mieux qualifier l'alternative qui s'offre à nous, rappelons que les algorithmes que nous considérons construisent un graphe représentant une distribution de probabilités donnée. Plus précisément, et ainsi que nous l'avons déjà indiqué dans la sous-section 1.2.1, ces algorithmes ont pour résultats des patrons causaux, dont chacun représente l'ensemble des graphes orientés acycliques compatibles (au sens de la définition 1.3) avec une distribution de probabilités donnée. Les flèches qui figurent dans un patron causal représentent donc celles des relations de cause à effet directes qu'il est possible d'inférer de la distribution de probabilités. En analysant les algorithmes, nous pourrons donc mettre au jour des conditions suffisantes de causalité.

Il nous semble que ce rappel nous donne deux raisons d'emprunter la voie

qui consiste à s'intéresser à la définition même des réseaux bayésiens causaux, plutôt que celle qui porte à analyser les algorithmes RB d'inférence aux causes. En premier lieu, il apparaît que la voie correspondant aux algorithmes RB implique une focalisation sur les conditions *suffisantes* pour « X cause Y ». Or, cette focalisation n'est pas imposée par la voie de l'analyse de la définition des réseaux bayésiens causaux et, par ailleurs, n'est pas requise *a priori* par le projet qui est le nôtre dans le présent chapitre. Il s'avérera d'ailleurs que c'est précisément sur le point des conditions *nécessaires* de causalité que la comparaison avec les théories probabilistes de la causalité est la plus instructive.

En second lieu, il nous semble qu'emprunter la voie qui consiste à s'intéresser directement aux algorithmes d'inférence causale risque de rendre délicate la distinction entre ce qui relève de l'analyse de la causalité sur laquelle l'inférence aux causes repose – et qui constitue l'objet spécifique de notre analyse – et ce qui revèle de la méthodologie de l'inférence causale. A l'inverse, la distinction sera maintenue sans effort si nous choisissons de revenir à la définition même des réseaux bayésiens. Dans un temps ultérieur, qui constitue la dernière sous-section du chapitre, nous nous pencherons sur les algorithmes pour montrer quel rôle jouent les éléments de caractérisation que nous nous apprêtons à mettre au jour relativement à la possibilité d'inférer des relations causales à partir de données probabilistes.

2.1.2 Mise au jour de la caractérisation RB

Ainsi que nous l'avons montré dans le chapitre 1, l'inférence aux causes fondée sur les réseaux bayésiens requiert que soient satisfaites trois hypothèses : l'hypothèse d'acyclicité, l'hypothèse de fidélité et la condition de Markov causale. De ces trois hypothèses, seules les deux dernières portent sur le rapport entre la causalité et les probabilités. C'est donc sur elles que doit se concentrer l'analyse si elle vise, comme c'est le cas ici, à mettre au jour la caractérisation de la causalité qui est à l'oeuvre dans les algorithmes RB d'inférence causale.

Avant et en vue de mettre cette caractérisation au jour, rappelons qu'un ensemble de variables $\mathbf{V}$:

1. satisfait la condition de Markov causale si et seulement si toute variable de $\mathbf{V}$ est indépendante de toutes les variables de $\mathbf{V}$ qui n'en sont pas des effets (directs dans $\mathbf{V}$ ou non) relativement à l'ensemble de ses causes directes dans $\mathbf{V}$;
2. satisfait l'hypothèse de fidélité si et seulement s'il n'existe pas dans $\mathbf{V}$ d'indépendances probabilistes relatives qui ne sont pas impliquées par

la condition de Markov causale.

La notion d'indépendance relative qui est mobilisée ici est définie dans l'appendice au chapitre 1 (définition 1.15).

D'après les rappels qui précèdent, si une variable X d'un ensemble $\mathbf{V}$ satisfaisant la condition de Markov causale ne cause pas une variable Y du même ensemble, alors X et Y sont indépendantes relativement à l'ensemble des causes directes de X dans $\mathbf{V}$. Par ailleurs, si $\mathbf{V}$ satisfait non seulement la condition de Markov causale, mais encore l'hypothèse de fidélité, alors *seules* les variables Y qui ne sont pas des effets de X sont indépendantes de X relativement à l'ensemble de ses causes directes dans $\mathbf{V}$. Dans ce cas, que X et Y soient indépendantes relativement à l'ensemble des causes directes de X dans $\mathbf{V}$ est à la fois une condition nécessaire et une condition suffisante pour que X ne cause pas Y. Par contraposition, on obtient le résultat suivant :

Théorème 2.1 (Condition nécessaire et suffisante de causalité) *Soit $\boldsymbol{V}$ un ensemble de variables qui satisfait la condition de Markov causale et l'hypothèse de fidélité, et soient X et Y deux variables de $\boldsymbol{V}$.*
X cause Y dans $\boldsymbol{V}$ si et seulement si X et Y sont dépendantes relativement à l'ensemble des causes directes de X dans $\boldsymbol{V}$.

Le résultat 2.1 suit immédiatement du double énoncé de la condition de Markov causale et de l'hypothèse de fidélité. Surtout, l'établir revient à reconnaître que les méthodes d'inférence aux causes fondées sur les réseaux bayésiens reposent sur la proposition suivante :

Proposition 2.1 (Caractérisation RB) *X cause Y dans $\boldsymbol{V}$ si et seulement si X et Y sont dépendantes relativement à l'ensemble des causes directes de X dans $\boldsymbol{V}$.*

C'est cette analyse de la causalité qu'il nous faut comparer avec les théories probabilistes. En vue de cette comparaison, nous consacrons la prochaine sous-section à mettre au jour une de ses conséquences.

2.1.3 Une conséquence de la caractérisation RB de la causalité

Dans le sous-section qui commence, nous montrons qu'il découle de la caractérisation RB de la causalité que la dépendance probabiliste relative à l'ensemble vide – c'est-à-dire la dépendance probabiliste absolue[1] – est une

[1]Dans la suite, nous ne mentionnerons plus cette précision quand elle est nécessaire. Ainsi, nous ne parlerons plus de dépendance (ou d'indépendance) absolue, mais plus simplement de dépendance (ou d'indépendance).

condition nécessaire de causalité. Pour le dire autrement, nous établissons le résultat suivant :

Théorème 2.2 (Condition nécessaire de causalité) *Soit* $\boldsymbol{V}$ *un ensemble de variables qui satisfait la condition de Markov causale et l'hypothèse de fidélité et soient* X *et* Y *deux variables de* $\boldsymbol{V}$*.*
Si X *cause* Y*, alors* X *et* Y *sont dépendantes relativement à l'ensemble vide.*

Pour établir ce résultat, le plus simple est de s'appuyer sur la notion de d-séparation que nous avons définie plus haut (définition 2.1.3) et sur le théorème 1.1 établi par Verma et Pearl. Une preuve est alors la suivante :

Preuve : Soit **V** un ensemble de variables qui satisfait la condition de Markov causale et l'hypothèse de fidélité, et soit X et Y deux variables de **V** telles que X cause Y.
On remarque que $\{X\}$ et $\{Y\}$ ne sont pas d-séparés par l'ensemble vide dans le graphe GC qui représente les relations de cause à effet directes dans **V**. En effet, le chemin c de GC qui correspond à ceci que X est une cause de Y ne satisfait aucune des deux conditions de d-séparation par l'ensemble vide qui découlent de la définition donnée dans le premier chapitre. Autrement dit, c n'est pas d-séparé par l'ensemble vide, d'où il découle que $\{X\}$ n'est pas d-séparé de $\{Y\}$ par l'ensemble vide dans GC.
Dans ces conditions, le théorème 1.1 implique qu'il existe une distribution de probabilités représentée par GC pour laquelle X et Y sont dépendants. Autrement dit, la satisfaction de la condition de Markov causale par **V** n'implique pas l'indépendance de X et Y.
Or, nous avons supposé que **V** satisfait non seulement la condition de Markov causale, mais encore l'hypothèse de fidélité. Il en découle que X et Y sont dépendants pour la distribution de probabilités dont il est question ici**.**

La condition nécessaire de causalité que nous venons de mettre au jour n'est pas suffisante. Néanmoins, il apparaîtra que l'avoir mise au jour est pertinent en vue de discuter des conditions de causalité qui sont à l'oeuvre dans les méthodes d'inférence aux causes fondées sur les réseaux bayésiens, et en particulier en vue de les comparer avec les conditions mises en avant par les théories probabilistes de la causalité. Toujours dans l'optique de cette comparaison, nous consacrons la prochaine section à présenter les théories probabilistes de la causalité, ou au moins ce qu'il est nécessaire d'en connaître pour que la comparaison avec la caractérisation RB soit à la fois possible et digne d'intérêt.

2.2 Théories probabilistes de la causalité

Les théories probabilistes de la causalité ne constituent pas l'objet de notre travail. Plus précisément, notre travail prend ces théories comme un point de départ qui n'est pas discuté pour lui-même. En conséquence, et ainsi que nous venons de le suggérer, la présentation que nous en donnons ici n'est pas exhaustive. Elle vise uniquement à interroger la caractérisation de la causalité qui est à l'oeuvre quand on infère des causes en utilisant les réseaux bayésiens. A cet effet, nous adoptons un mode de présentation d'inspiration historique, que nous reprenons pour partie de Hitchcock (2002). Ce mode de présentation consiste à montrer comment, à partir de l'idée séminale selon laquelle une cause augmente la probabilité de son effet, les théories probabilistes successives visent à prendre en charge un nombre de plus en plus élevé de classes de contre-exemples à l'analyse proposée.

2.2.1 L'idée séminale

Ainsi que nous venons de le rappeler, l'idée sur laquelle les théories probabilistes de la causalité sont fondées est celle de caractériser une cause par ceci qu'elle augmente la probabilité de ses effets. Rappelons que cette idée séminale se spécifie de la façon suivante :

Analyse probabiliste de la causalité 2.1 (Idée séminale) *A cause B si et seulement si $p(B \mid A) > p(B)$.*

Ici, et conformément à ce que nous avons plus haut dit de la causalité générique, A et B sont des propriétés. Notons par ailleurs que cette proposition de caractérisation de la causalité n'a de sens que si la probabilité de A est non nulle. Ici et dans la suite, nous le supposons sans le mentionner. Cette hypothèse n'est guère problématique dès lors qu'il n'y a pas de sens immédiatement intelligible à parler d'une cause dont la probabilité est nulle.

L'énoncé de l'idée qu'on trouve au fondement des théories probabilistes de la causalité est utilement complétée par la proposition suivante :

Proposition 2.2 *Etant données deux propriétés A et B, les deux propositions suivantes sont équivalentes :*

1. $p(B \mid A) > p(B)$
2. $p(B \mid A) > p(B \mid \text{non-}A)$.

Nous en donnons la démonstration suivante :

Preuve : Soient A et B deux propositions.
Par le théorème des probabilités totales :

p(B) = p(B | A).p(A) + p(B | non-A).p(non-A).
Il en découle que :
p(B | A) > p(B | non-A) équivaut à p(B) < p(B | A).p(A) + p(B | A).p(non-A).
Or :
p(B | A).p(A) + p(B | A).p(non-A) = p(B | A).[p(A) + p(non-A)]
p(B | A).p(A) + p(B | A).p(non-A) = p(B | A).[p(A) + 1 - p(A)] = p(B | A).
Dès lors, p(B | A) > p(B | non-A) équivaut à p(B | A) > p(B)**.**

De la proposition que nous venons d'établir, il découle que l'idée selon laquelle A augmente la probabilité de B est exprimée par l'inégalité « p(B | A) > p(B | non-A) », aussi bien que par « p(B | A) > p(B) ».

Avant d'en montrer les limites, il convient d'insister sur le caractère plausible de l'idée qu'on trouve au fondement des théories probabilistes de la causalité. En effet, il semble vrai que la probabilité d'une propriété est plus élevée si l'une de ses causes est présente, que si elle ne l'est pas. Pour reprendre la formulation proposée par Cartwright, la raison en est que « les causes produisent leurs effets ; elles les font advenir »[2]. A titre d'illustration, dire que fumer cause le cancer, c'est bien en particulier dire que la probabilité de développer un cancer est plus élevée pour les fumeurs qu'elle ne l'est pour les non-fumeurs.

Il reste que l'augmentation de probabilité ne suffit pas à caractériser une cause. Plus précisément, l'augmentation de probabilité ne suffit pas à caractériser cet aspect de la causalité que les théories probabilistes visent à capturer. Ainsi, il existe des cas d'augmentation de probabilité qui ne sont pas à mettre au compte d'une relation de cause à effet. On parlera alors de « corrélations trompeuses » (*spurious correlations*). L'existence de corrélations trompeuses n'a pas échappé aux tenants des théories probabilistes de la causalité, qui n'ont jamais prétendu faire de l'augmentation de probabilité une condition nécessaire et suffisante de causalité. Il nous revient maintenant de montrer comment ils ont tenu compte de la possibilité que des corrélations soient trompeuses.

2.2.2 Deux types de corrélations trompeuses

Parmi les situations dans lesquelles une propriété augmente la probabilité d'une propriété qu'elle ne cause pas, deux types sont aisément repérables et sont déjà pris en compte dans les premières théories probabilistes de la causalité. Nous les présentons ici.

[2]Cartwright (2001) p. 255.

2.2.2.1 Corrélations entre effets et causes

Les corrélations du premier des deux types que nous présentons ici sont engendrées par ceci que si une cause augmente la probabilité de son effet, alors un effet augmente la probabilité de sa cause. En effet, la relation d'augmentation de probabilité est symétrique :

Proposition 2.3 *Soient A et B deux propriétés de probabilité non nulle. p(B | A) > p(B) si et seulement si p(A | B) > p(A).*

Cette proposition s'établit simplement :

Preuve : Soient A et B deux propriétés de probabilité non nulle.
Les propositions suivantes sont équivalentes :

- p(B | A) > p(B)
- p(B et A) / p(A) > p(B)
- p(B et A) / p(B) > p(A)
- p(A | B) > p(A).

Ainsi, en particulier, p(B | A) > p(B) équivaut à p(A | B) > p(A)**.**

De la proposition 2.3, il découle que l'augmentation de probabilité considérée comme condition nécessaire de causalité ne permet pas de distinguer entre les causes et les effets. Plus précisément, si une cause augmente la probabilité de ses effets, alors chacun de ces effets augmentent la probabilité de cette cause. Comme la relation de causalité n'est pas, de son côté, symétrique, il en résulte un premier type de corrélations trompeuses.

2.2.2.2 Corrélations entre effets d'une même cause

Le second type de corrélations trompeuses aisément repérable et tôt identifié dans l'histoire des théories probabilistes de la causalité est celui des corrélations entre effets d'une même cause. A titre d'illustration, revenons à la propriété d'être fumeur. Nous avons déjà mentionné à plusieurs reprises qu'elle cause la propriété de développer un cancer du poumon. D'un autre côté, elle cause la propriété d'avoir les doigts jaunis. Les deux propriétés de développer un cancer du poumon et d'avoir les doigts jaunis sont toutes deux produites, et donc rendues plus probables, par la propriété d'être fumeur. Il est alors vraisemblable que chacune de ces propriétés augmente la probabilité de l'autre : il est plus probable de développer un cancer du poumon quand on a les doigts jaunis que dans le cas général. Pourtant, on accordera que la propriété de développer un cancer du poumon ne cause pas la propriété d'avoir les doigts jaunis. Nous avons alors affaire, à nouveau, à une corrélation trompeuse.

Depuis les années 1950 au moins, on sait caractériser les situations dans lesquelles existent des corrélations trompeuses appartenant à ce second type.

Plus précisément, l'analyse de la direction du temps au moyen de la notion de fourche conjonctive dans Reichenbach (1956) repose sur l'identification d'une propriété des causes communes à plusieurs effets : quand on conditionnalise sur une telle cause, la dépendance probabiliste entre ses effets qui découle de ce qu'ils sont effets de cette même cause disparaît. Pour le dire autrement, deux propriétés A et B qui dépendantes l'une de l'autre parce qu'elles sont effets d'une même cause C sont indépendantes relativement à cette cause : p(B | A et C) = p(B | C). Ainsi, relativement à la propriété d'être fumeur, les propriétés de souffrir d'un cancer du poumon et d'avoir les doigts jaunis sont indépendantes en probabilité. En effet, une fois la propriété d'être fumeur prise en compte, avoir les doigts jaunis ne change plus rien à la probabilité de souffrir d'un cancer du poumon. Dans des termes introduits par Reichenbach et qui continuent d'être utilisés aujourd'hui, la propriété d'être fumeur fait écran entre (*screen off*) les propriétés de souffrir du cancer du poumon et d'avoir les doigts jaunis.

2.2.2.3 La théorie de Suppes

Dans Suppes (1970), Suppes s'appuie sur la propriété des causes communes de faire écran entre leurs effets pour proposer l'une des premières théories probabilistes de la causalité.[3] Plus précisément, il utilise cette propriété des causes communes pour caractériser (en termes probabilistes) les corrélations trompeuses du second type, puis il adjoint à l'idée séminale d'augmentation de probabilité une clause stipulant que les corrélations trompeuses du second type ne correspondent pas à des relations de cause à effet.

Le cas des corrélations trompeuses entre un effet et ses causes reçoit, quant à lui, un traitement temporel : une cause et l'un de ses effets se distinguent par ceci que la première est antérieure au second. La théorie alors obtenue par Suppes peut être énoncée dans les termes suivants :

Analyse probabiliste de la causalité 2.2 (Suppes, 1970) *A cause B si et seulement si :*

1. *A est antérieur à B*
2. *p(B | A) > p(B)*
3. *il n'existe pas de C antérieur à A tel que p(B | A et C) = p(B | C).*

[3]La théorie développée dans Good (1961) et Good (1962) n'a pas eu l'impact de celle de Suppes et reste largement ignorée de l'histoire usuelle des théories probabilistes de la causalité. Surtout, cette histoire usuelle nous suffit pour mettre en perspective la caractérisation RB de la causalité en la comparant aux théories probabilistes. Aussi faisons-nous commencer à Suppes (1970) notre présentation des théories probabilistes de la causalité.

Dans le cas où les deux premières clauses sont satisfaites, Suppes parle de « cause *prima facie* ». Selon Suppes, ses théories de la causalité et de la causalité *prima facie* valent aussi bien de la causalité singulière que de la causalité générique. Pour des raisons que nous avons déjà données, c'est seulement en tant que théorie probabiliste de la causalité *générique* que nous envisageons ici sa théorie de la causalité.

2.2.3 Limites de la théorie de Suppes

La théorie de la causalité générique qui est développée dans Suppes (1970) connaît cinq limites principales, que nous présentons maintenant.

2.2.3.1 Corrélations trompeuses entre effets d'une même cause *interactive*

La théorie proposée par Suppes se heurte d'abord à ceci que sa clause 3. ne suffit pas à identifier toutes les corrélations trompeuses entre effets d'une même cause. Pour le dire autrement, il existe des effets d'une même cause qui sont dépendants seulement parce qu'ils sont effets de cette même cause, mais entre lesquels cette cause ne fait pourtant pas écran. C'est le cas quand cette cause et ses deux effets composent une fourche interactive (au sens défini dans le paragraphe 1.3.2.1). Dans ce cas, nous dirons de la cause elle-même qu'elle est interactive.

Nous avons vu dans le paragraphe 1.3.2.1 qu'un exemple canonique de fourche interactive est celui qui est proposé dans Cartwright (1999) : une usine qui ne produit pas à chaque fois qu'on lui en donne l'ordre, mais qui pollue exactement quand elle produit. Pour ce qui est des théories probabilistes de la causalité, cet exemple est pertinent en tant que conditionnaliser sur la cause ne suffit pas à rendre ses effets indépendants. Par ailleurs, on ne voit pas qu'il existe un C différent de la cause commune qui fasse écran entre les effets. Dans les situations de ce type, la théorie de Suppes conduit à tort à considérer que les deux effets de la cause interactive sont dans une relation de cause à effet. C'est la première des limites de cette théorie. Nous retiendrons de ce premier paragraphe l'idée générique selon laquelle la théorie de Suppes n'est pas adéquate pour les causes interactives. Dès lors, nous ne traitons dans la suite que des causes qui ne le sont pas, et des limites rencontrées par la théorie de Suppes dans les contextes que nous qualifierons de « non interactifs ».

2.2.3.2 Traitement des corrélations entre effets et causes

Pas plus qu'elle ne résout complètement le problème des corrélations trompeuses entre effets d'une même cause, la théorie de Suppes ne résout correctement le problème des corrélations trompeuses entre les effets et leurs causes. Pour le dire autrement, le critère proposé par Suppes pour le sens des relations de cause à effet est insatisfaisant. Il l'est précisément à deux titres distincts. D'une part, il suppose qu'on peut ordonner temporellement des propriétés. Or, il n'est pas clair comment cela peut être fait. On pourrait envisager de le faire en se référant aux relations de cause à effet singulières qui correspondent à la relation de cause à effet générique considérée. En effet, tels que nous les avons définis dans l'introduction à notre travail, les *relata* de la causalité singulière sont temporellement situés et, du coup, ordonnables.

Toutefois, et quand bien même on réussirait à ordonner temporellement des propriétés, la solution proposée par Suppes au problème du sens de la causalité suppose d'autre part que la causalité générique est asymétrique. Or, nous avons défendu dans le premier chapitre de notre travail la thèse selon laquelle elle ne l'est pas. Dans ces conditions, la clause 1. de la théorie proposée n'est pas convenable. Positivement, une théorie probabiliste de la causalité générique n'est satisfaisante que si elle comporte les moyens de distinguer entre une cause et son effet, mais n'impose pas à la relation de causalité générique l'asymétrie dont nous avons vu (dans le paragraphe 1.3.1.1) qu'elle n'est pas la sienne.

Nous venons de voir que la théorie de Suppes ne traite pas correctement les cas de corrélations trompeuses entre effets et causes d'abord, et entre effets d'une même cause ensuite. En d'autres termes, elle ne traite pas correctement tous les cas de corrélations trompeuses qu'elle prend en compte – c'est-à-dire, précisément, qu'elle vise à traiter. Une autre difficulté, sensiblement différente, est qu'il existe des corrélations trompeuses qui ne relèvent ni de l'un, ni de l'autre des deux types que la théorie de Suppes vise à prendre en compte. Ces corrélations ne sont même pas prises compte dans le cadre de la théorie de Suppes. *A fortiori* elles ne sont pas traitées correctement par elle. En nous appuyant sur la fin de Cartwright (2001)[4], nous considérons que les corrélations trompeuses non prises en compte dans le cadre de la théorie de Suppes relèvent de deux grands types. Nous présentons ces deux types dans les deux paragraphes à venir. En vue de cette présentation, nous nous rappelons au lecteur que les corrélations trompeuses qui nous intéressent sont des dépendances probabilistes absolues et qui se rencontrent dans les contextes non interactifs.

[4]Cartwright (2001) p. 254 et suivantes.

2.2.3.3 Corrélations trompeuses entre effets de plusieurs causes

En premier lieu, la théorie de la causalité proposée par Suppes ne tient pas compte de la possibilité qu'existent des corrélations trompeuses entre effets de *plusieurs* causes.

Ainsi, elle ne tient pas compte de l'existence de corrélations trompeuses entre des effets *communs* à plusieurs causes, qui sont engendrées par cette communauté de causes, et telles qu'aucune de ces causes ne suffit à faire écran entre les effets. Pour qu'une corrélation entre A et B qui est due à des causes communes soit identifiée comme trompeuse dans le cadre de la théorie de Suppes, il faut qu'*une* cause commune suffise à faire écran entre A et B. Dès lors que ce n'est pas le cas, et en particulier même si une conjonction de causes communes fait écran entre A et B, la théorie de Suppes implique à tort qu'il existe entre A et B une relation de cause à effet.

La théorie de Suppes ne tient pas compte non plus de l'existence de corrélations entre des propriétés qui d'une part ont plusieurs causes et d'autre part n'entretiennent aucune forme de rapport causal – c'est-à-dire qu'aucune n'est cause de l'autre et qu'elles n'ont pas de cause en commun. Pour comprendre ce qui est en jeu ici, revenons à l'exemple introduit dans Sober (1988), celui de la corrélation entre le prix du pain en Grande-Bretagne et le niveau des mers à Venise. Pour retrouver le langage des propriétés, envisageons la corrélation – supposée trompeuse – entre la propriété du prix du pain britannique d'être élevé et la propriété du niveau des mers à Venise d'être important. Dans un premier temps, imaginons que le prix du pain en Grande-Bretagne dépende causalement du seul prix du blé en Grande-Bretagne. Dans ce cas, conditionnaliser sur le prix du blé suffit à faire écran entre le prix élévé du pain en Grande-Bretagne et le niveau important des mers à Venise.[5] Imaginons maintenant que le prix du pain en Grande-Bretagne dépende à la fois du prix du blé et de la demande de pain. Alors, il n'existe pas une cause C qui suffise à faire écran entre les propriétés du prix du pain britannique d'être élevé et du niveau des mers à Venise d'être important. Dès lors, la théorie de Suppes implique à tort que la propriété du prix du pain britannique d'être élevé cause la propriété du niveau des mers à Venise d'être important.

De façon générale, il semble clair qu'une corrélation entre deux propriétés A et B qui n'entretiennent aucun rapport causal n'est pas identifiée comme trompeuse dans le cadre de la théorie de Suppes si *plusieurs* propriétés sont nécessaires pour faire écran entre elles. Le problème, ici, est que la clause 3. de la théorie de Suppes exige qu'*une* propriété C suffise à faire écran entre deux propriétés pour que leur corrélation soit reconnue comme trompeuse. Il

[5] On notera au passage qu'on met ici au jour un type de corrélations trompeuses que la théorie de Suppes prend en charge alors même qu'elle ne le vise pas explicitement.

en découle que la théorie de Suppes échoue donc à traiter correctement des corrélations trompeuses entre les effets de *plusieurs* causes.

2.2.3.4 Le paradoxe de Simpson

En second lieu, la théorie proposée par Suppes ne prend pas en compte les corrélations trompeuses qui participent du phénomène connu par les statisticiens sous le nom de « paradoxe de Simpson ». A proprement parler, le paradoxe de Simpson consiste en ceci que certaines corrélations positives (resp. négatives) peuvent devenir négatives (resp. positives) quand on s'intéresse à des sous-populations de la population initialement envisagée. Dans le langage des propriétés : il existe des corrélations positives (resp. négatives) entre propriétés qui deviennent négatives (resp. positives) quand on conditionnalise sur la propriété d'appartenir à une – n'importe laquelle – des sous-populations, pour une partition donnée, de la population initialement considérée.

A titre d'illustration, venons en à l'exemple le plus fameux pour ce qui est du paradoxe de Simpson. Cet exemple est relatif à l'admission en troisième cycle (*graduate studies*) à l'université de Berkeley. De façon générale, le taux d'admission est sensiblement moins élevé pour les filles que pour les garçons. Ainsi pour l'année 1973, 35 % des candidates contre 44 % des candidats ont été admis en troisième cycle. Mais, cette même année, le pourcentage de filles admises est plus élevé que le pourcentage de garçons admis dans presque tous les départements de l'université considérés isolément.[6] En d'autres termes, pour presque tous les départements de l'université de Berkeley, conditionnaliser sur la propriété de se porter candidat dans tel département, inverse le sens de la dépendance probabiliste entre la propriété d'être une fille et la propriéte d'être admis en troisième cycle. Notons que « presque » n'est pas essentiel d'un point de vue mathématique : on pourrait construire un cas (fictif) dans lequel le pourcentage d'admission des garçons reste plus élevé que celui des filles au niveau global, mais est moins élevé que lui dans *tous* les départements.[7]

L'exemple de l'admission en troisième cycle à Berkeley nous intéresse ici pour la raison suivante : sous l'hypothèse selon laquelle les fréquences observées sont de bons indicateurs des probabilités, la théorie de la causalité que propose Suppes conduit à conclure que la propriété d'être une fille cause la propriété de ne pas être admis. En effet, la propriété d'être une fille augmente la probabilité de ne pas être admis et il n'existe pas de propriété qui

[6]Pour les chiffres exacts, on se reportera à l'article dans lequel le cas est introduit : Bickel et al. (1975).

[7]Ce qui joue en fait est que les filles se portent massivement candidates dans les départements pour lesquels les taux d'admission sont les plus faibles.

fasse écran entre les deux premières. Pourtant, de l'autre côté, les résultats obtenus en conditionnalisant sur la propriété de se porter candidat dans tel ou tel département suggèrent que cette conclusion n'est pas satisfaisante. Positivement, ces résultats suggèrent que la relation de cause à effet n'est pas entre les propriétés d'être une fille et de ne pas être admis en troisième cycle à Berkeley, mais bien plutôt entre celles d'être une fille et d'être admis en troisième cycle à Berkeley.

Le paradoxe de Simpson engage des corrélations trompeuses différentes de toutes celles que nous avons envisagées jusqu'à présent. Ainsi, dans les instances du paradoxe, les corrélations ne sont pas trompeuses parce qu'elles ne correspondent pas à des relations de cause à effet, mais parce qu'elles sont de mauvais indicateurs des propriétés qui sont en relation de cause à effet. Mais de même que les corrélations trompeuses envisagées dans les paragraphes précédents, celles qu'engage le paradoxe de Simpson ne sont pas prises en compte par la théorie de Suppes. En effet, la théorie implique à tort que les corrélations repérées dans la population initiale correspondent à des relations de cause à effet ou – s'il existe une sous-population dans laquelle les corrélations s'annulent – qu'elles ne correspondent à aucune relation de cause à effet. Dans aucun des deux cas elle ne permet d'identifier correctement la relation de cause à effet entre la propriété d'être une fille et celle d'être admis en troisième cycle à Berkeley. Le paradoxe de Simpson occasionne donc bien un quatrième type de contre-exemples à la théorie de la causalité proposée par Suppes. Celui-ci étant mis au jour, nous pouvons en venir au cinquième et dernier des types de contre-exemples à cette théorie.

2.2.3.5 Indépendances trompeuses

Dans le paragraphe précédent, nous avons indiqué qu'il est possible qu'une propriété A diminue la probabilité d'une propriété B dans une population P, mais l'augmente dans toutes les sous-populations de P pour une partition donnée. C'est le cas de la propriété d'être une fille relativement à la propriété d'être admis en troisième cycle à Berkeley. Maintenant et de façon analogue, il est possible que A laisse la probabilité de B inchangée dans P, mais l'augmente dans toutes les sous-populations de P pour une partition donnée. A titre d'illustration, on peut imaginer qu'il existe entre les propriétés d'être fumeur et de pratiquer un exercice physique régulier une corrélation telle que, dans la population étudiée, la propriété de fumer devienne indépendante de son effet qu'est la propriété de souffrir de problèmes cardiaques. Pour dire les choses plus clairement, on peut imaginer que l'effet néfaste du tabac sur la santé cardiaque soit exactement contre-balancé par celui de l'activité physique régulière à laquelle la propriété de fumer serait par

hypothèse corrélée. Il n'en resterait pas moins que fumer cause la propriété de souffrir de problèmes cardiaques, et en augmente la probabilité à la fois parmi ceux qui pratiquent une activité physique régulière et parmi ceux qui n'en pratiquent pas. Cette relation de cause à effet n'est pas mieux identifiée par la théorie de Suppes que la relation entre la propriété d'être une fille et celle d'être admis en troisième cycle à Berkeley.

La difficulté, toutefois, est différente dans l'un et l'autre cas. Pour ce qui est de l'admission en troisième cycle à Berkeley, nous avons vu que le problème est que la corrélation est trompeuse – en un sens précis que nous avons défini – mais que la clause 3. de la théorie de Suppes ne suffit pas à l'identifier comme telle. Dans le cas qui nous occupe maintenant, le problème est plutôt que la relation de cause à effet n'est signalée en première instance par aucune dépendance probabiliste (trompeuse ou non) et que, du coup, la clause 2. de la théorie de Suppes implique qu'il n'y pas de relation de cause à effet. Pour dire les choses autrement, il vient d'apparaître successivement deux choses. En premier lieu, il existe non seulement des corrélations, mais encore des *indépendances* trompeuses. En second lieu, l'existence de ces indépendances trompeuses n'est pas prise en compte par la théorie de Suppes, dont la clause 2. fait de l'augmentation de probabilité une condition nécessaire pour qu'il y ait relation de cause à effet.

Avant de montrer comment les indépendances trompeuses sont traitées dans le cadre des théories probabilistes de la causalité postérieures à celle de Suppes, il convient d'indiquer que les indépendances trompeuses ne sont pas toutes du type de celle que nous venons d'envisager. Il existe des situations plus complexes qui elles aussi donnent lieu à des indépendances trompeuses. Il apparaîtra dans la suite que la mise en perspective de la caractérisation RB de la causalité n'impose pas de décrire ces situations. Pour une typologie et une analyse détaillée, nous renvoyons le lecteur à Cartwright (1989) ; pour un exemple classique, nous le renvoyons à Hesslow (1976).

2.2.4 Théories probabilistes de la causalité après Suppes (1970)

Les théories probabilistes de la causalité qui succèdent à celle de Suppes visent d'abord à prendre en considération les indépendances trompeuses du type de celle que nous avons envisagée et les corrélations trompeuses qui composent les instances du paradoxe de Simpson. Dans les deux cas, la difficulté pour la théorie de Suppes procède de ce qu'elle fait de l'augmentation de probabilité dans une population, une condition nécessaire pour la causalité dans cette population.

En même temps, les deux cas suggèrent que la causalité peut bien être analysée comme une augmentation de probabilité s'il ne s'agit pas de probabilité dans la population considérée, mais dans certaines de ses sous-populations. Dans l'exemple de l'indépendance trompeuse entre la propriété d'être fumeur et celle de souffrir de problèmes cardiaques, il s'agit d'une part de la sous-population de ceux qui pratiquent une activité physique régulière et d'autre part la sous-population de ceux qui n'en pratiquent pas. Pour l'admission en troisième cycle à Berkeley, ces sous-populations sont définies par le département de candidature. Dans les deux cas, les sous-populations pertinentes pour la causalité sont caractérisées par une propriété : dans le premier cas celles de pratiquer et de ne pas pratiquer une activité physique régulière, dans le second cas celle de se porter candidat dans tel ou tel département. Dès lors, un moyen de considérer les probabilités dans l'une de ces sous-populations plutôt que dans la population tout entière est de conditionaliser sur la propriété qui la caractérise. L'idée qui se fait jour alors est de caractériser la causalité non pas par l'augmentation de la probabilité absolue, mais par l'augmentation des probabilités conditionnelles relatives à ces sous-populations. Pour le dire autrement, il s'agirait de ne pas considérer des relations du type $p(E \mid C) > p(E)$, mais plutôt des relations du type $p(E \mid C.A) > p(E \mid A)$ avec A caractérisant une sous-population.

Au-delà de ces cas particuliers, il reste à donner une définition générale des sous-populations qu'il faut considérer, et donc des propriétés sur lesquelles il faut conditionnaliser, pour pouvoir proposer une théorie de la causalité comme augmentation de probabilité. En vue de cela, on peut revenir aux deux exemples que nous avons proposés. On remarquera alors que les sous-populations à définir sont caractérisées par leur homogénéité relativement à des propriétés différentes de la cause envisagée, mais qui comme elle peuvent avoir une influence sur l'effet. Ainsi la propriété de pratiquer régulièrement une activité physique influence-t-elle celle de souffrir de problèmes cardiaques. De la même façon à Berkeley, se porter candidat dans tel département plutôt que dans tel autre influence l'admission en trois cycles. L'influence en question est, dans les deux cas, causale.

Il apparaît alors que les sous-populations que nous cherchons à définir sont des sous-populations d'individus homogènes relativement à toutes les causes de l'effet considéré qui diffèrent de la cause considérée. Aussi, l'idée qui point ici est celle d'analyser la causalité comme une augmentation de probabilité toutes choses étant égales par ailleurs et de considérer que toutes choses sont égales par ailleurs quand est fixée la situation relativement aux causes différentes de celle qui fait l'objet de l'analyse. En suivant cette ligne d'analyse, on aboutit à l'idée développée dans Cartwright (1979) :

C cause E si et seulement si C augmente la probabilité de E dans

> toute situation qui est par ailleurs (*otherwise*) causalement homogène par rapport à E.[8]

En termes plus formels, la proposition est la suivante :

Analyse probabiliste de la causalité 2.3 (Cartwright, 1979) *A cause B si et seulement si* $p(B \mid A.\boldsymbol{S}_i) > p(B \mid \boldsymbol{S}_i)$ *pour tout* $\boldsymbol{S}_i$*, où les* $\boldsymbol{S}_i$ *sont les descriptions d'états sur l'ensemble des propriétés qui causent B mais ne sont pas causées par A.*

Sous cette définition des $\mathbf{S}_i$, chacun d'entre eux décrit bien une sous-population de la population considérée qui est minimale parmi celles qui sont homogènes relativement aux causes de B qui diffèrent de A. La précision selon laquelle ces propriétés ne sont pas elles-mêmes causées par A est rendue nécessaire par ceci que les intermédiaires causaux font écran à la dépendance probabiliste entre une cause et ses effets : en règle générale, en conditionalisant sur une cause de B qui est causée par A, on rend A et B indépendantes en probabilité. D'une cause de B qui n'est pas causée par A, on dira qu'elle est « indépendante de A ».

La théorie 2.3 que nous venons de présenter n'est pas la seule qui repose sur une analyse du type de celle que nous avons proposée pour les quatrième et cinquième limites rencontrées par la théorie de Suppes. Plus explicitement, Skyrms (1980) contient une théorie probabiliste de la causalité proche de celle qui est développée dans Cartwright (1979). Elle en diffère par ceci que Skyrms considère que la causalité de A à B ne requiert pas que A augmente la probabilité de B dans *toutes* les situations homogènes relativement aux causes de B indépendantes de A. Selon Skyrms, pour que A cause B, il suffit que A augmente la probabilité de B dans *une* situation homogène relativement aux causes de B indépendantes de A et ne la diminue dans aucune. La théorie alors proposée est la suivante :

Analyse probabiliste de la causalité 2.4 *A cause B si et seulement si :*

1. *il existe j tel que* $p(B \mid A.\boldsymbol{S}_j) > p(B \mid \boldsymbol{S}_j)$ *;*
2. *il n'existe pas k tel que* $p(B \mid A.\boldsymbol{S}_k) < p(B \mid \boldsymbol{S}_k)$

où les $\boldsymbol{S}_i$ *sont les descriptions d'états sur l'ensemble des causes de B indépendantes de A.*

On notera immédiatement que la forme qu'elles partagent confère aux théories de Cartwright et de Skyrms (au moins) une caractéristique remarquable : toutes deux sont ce que nous avons appelé des analyses « circulaires » de la causalité. Dans les deux cas, en effet, il apparaît des concepts causaux

[8] Cartwright (1979) p. 423.

– et plus précisément le concept même de cause – dans l'*analysans* proposé pour « A cause B ».

Pour ce qui est, maintenant, de la pertinence de l'analyse proposée, nous avons indiqué déjà que les théories 2.3 et 2.4 sont construites pour prendre en charge les instances du paradoxe de Simpson et celles des indépendances trompeuses qui leur sont structurellement similaires. Restent maintenant les autres difficultés auxquelles la théorie de Suppes se heurte. Concernant, d'abord, la distinction entre cause et effet (dont nous avons montré dans le paragraphe 2.2.3.2 qu'elle est problématique chez Suppes), elle est bien assurée dans le cadre de ces théories. D'un côté, en effet, l'analyse proposée n'est pas symétrique : là où l'analyse de « A cause B » mobilise la notion de cause de B indépendante de A, celle de « B cause A » mobilise la notion de cause de A indépendante de B. De l'autre côté, les théories 2.3 et 2.4 n'interdisent pas que deux propriétés A et B se causent mutuellement.

Concernant, ensuite, les corrélations entre effets de plusieurs causes qui sont trompeuses mais que la théorie de Suppes échouent à identifier comme telles (voir le paragraphe 2.2.3.3), elles sont correctement analysées dans le cadre des théories de Cartwright et de Skyrms. En effet, ces deux analyses requièrent pour « A cause B » une augmentation de probabilité alors que la situation est homogène relativement à *toutes* les causes de B indépendantes de A. Pour qu'une corrélation qui ne correspond pas à une relation de cause à effet soit correctement caractérisée comme trompeuse, il n'est donc plus nécessaire qu'*une* cause suffise à lui faire écran – ce qui était exactement le point problématique dans le cadre de la théorie de Suppes.

D'un autre côté, les théories proposées par Cartwright et par Skyrms, pas plus que celle de Suppes, ne permettent d'identifier comme trompeuses les corrélations entre effets communs à une cause interactive . En effet, ce qui caractérise les causes interactives est précisément qu'elles ne font pas écran entre leurs effets. En outre, les théories de Cartwright et de Skyrms achoppent sur des indépendances trompeuses correspondant à des situations causales plus complexes que celle que nous avons explicitement envisagée dans le paragraphe 2.2.3.5 ci-dessus.

Traiter correctement ces situations a motivé l'introduction de théories probabilistes de la causalité plus raffinées que celles qui sont développées dans Cartwright (1979) et Skyrms (1980). Ces théories sont développées en particulier dans Eells et Sober (1983) et dans Cartwright (1989). Pas plus que celui des situations qui motivent leur introduction, le détail de ces théories ne nous intéressera ici. Il nous suffira d'en dire que, de même que les théories développées dans Cartwright (1979) et Skyrms (1980), elles reposent sur la description des contextes dans lesquels une cause augmente la probabilité de ses effets. De même que dans Cartwright (1979) et Skyrms (1980), cette

description requiert des concepts causaux. Autrement dit, au même titre que les théories développées dans Cartwright (1979) et Skyrms (1980) et pour les mêmes raisons, ces théories constituent des analyses circulaires de la causalité.

Notre présentation des théories probabilistes de la causalité s'achève ici. Malgré ses lacunes – en particulier pour ce qui est de la fin de l'histoire –, elle suffit à comparer ces théories avec la caractérisation RB de la causalité. C'est la tâche à laquelle nous nous attelons maintenant.

2.3 Caractérisation RB et théories probabilistes de la causalité

Si la caractérisation RB et les théories probabilistes de la causalité ont un air de famille, elles ne sont pas immédiatement commensurables. En effet, il semble clair qu'elles n'ont pas exactement le même objet. En particulier, on aura remarqué que les relations de cause à effet visées par la première sont relatives à un ensemble de variables, tandis que les secondes analysent les relations de cause à effet en général, dans l'absolu. Dès lors, un premier moment de la section qui commence est consacré à comparer les objets des deux types d'analyses que nous comparons. Dans le même temps, cette première sous-section nous permet de formuler la question de la comparaison des analyses dans des termes qui nous permettent d'y apporter une réponse. Nous apportons cette réponse dans la deuxième sous-section. Plus précisément, cette deuxième sous-section vise à situer la caractérisation RB de la causalité dans le portrait que nous venons de brosser de la famille des théories probabilistes de la causalité. Les conséquences pour l'inférence aux causes de ce que nous aurons montré alors sont explorées dans une troisième et dernière sous-section.

2.3.1 Comparaison des objets

Les objets de la caractérisation RB d'une part et des théories probabilistes de la causalité d'autre part diffèrent de deux manières significatives. Nous consacrons à chacune un paragraphe dans lequel nous la présentons, la discutons et travaillons à la réduire en vue de comparer non plus les objets des deux types d'analyses en présence, mais ces deux types d'analyses eux-mêmes. Ainsi pouvons-nous finalement énoncer la question de la comparaison entre ces deux types d'analyses dans des termes intelligibles et tels qu'elle pourra recevoir une réponse dans la suite de la section.

2.3.1.1 Causalité relative à un ensemble de variables et causalité absolue

La première des deux différences que nous discutons est celle que nous avons mentionnée dans l'introduction à la présente section. Plus explicitement, le paragraphe qui commence traite de ceci que les réseaux bayésiens permettent d'inférer des relations de cause à effet relatives à un ensemble de variables[9] là où les théories probabilistes visent à analyser les relations de cause à effet – tout court.

Les voies d'une réduction de cette première différence sont indiquées par Spohn :

> Si la notion de dépendance causale se présente d'abord comme relative à un cadre (*frame relative*), nous pouvons éliminer ce caractère relatif seulement en en venant au cadre qui embrasse tout (*all-embracive frame*), contenant toutes les variables nécessaires pour une description complète de la réalité empirique.[10]

Ainsi, la première différence entre l'objet de la caractérisation RB et celui des théories probabilistes se réduit sous l'hypothèse selon laquelle les relations de cause à effet absolues sont les relations de cause à effet relatives à un ensemble de variables qui suffit à décrire la réalité empirique. Cette hypothèse nous semble nécessaire pour qu'il y ait même un sens à s'intéresser aux relations de cause à effet relatives à un ensemble de variables – et donc aux réseaux bayésiens causaux. Aussi l'émettons-nous.

Maintenant, il peut sembler que la notion de relation de cause à effet relatives à un ensemble de variables qui suffit à décrire la réalité empirique n'est pas tenable. Par là, on entendrait à la fois que la notion est peu manipulable et que, à la limite, elle n'aurait pas de sens. A cette double objection, nous apportons deux éléments de réponse distincts. En premier lieu, il convient de rappeler que les théories probabilistes de la causalité visent une notion absolue de causalité. Considérer que ces théories sont dignes d'intérêt est donc admettre qu'une notion de ce type fait sens. Si nous ne l'admettions pas, notre travail n'aurait pas de sens.

En second lieu, notons que la manipulation effective des concepts causaux ne requiert pas à tout moment de prendre en compte toutes les variables d'un ensemble qui suffit à décrire la réalité empirique. Etant donné un phénomène, analyser sa structure causale ne requiert pas de prendre en compte toutes les variables d'un ensemble qui suffit à décrire la réalité empirique. Il suffit, et

[9] Un coup d'oeil jeté à la proposition 2.1 suffira à en convaincre le lecteur qui ne le serait pas déjà.

[10] Spohn (2001) p. 11.

il est même préférable, de considérer un sous-ensemble de cet ensemble qui soit tel que connaître les relations causales entre les variables de ce sous-ensemble permet de comprendre le phénomène. De cette façon, la notion de relation de cause à effet entre toutes les variables d'un ensemble qui suffit à décrire la réalité empirique devient manipulable en pratique, de même que les théories probabilistes de la causalité. Nous considérerons donc dans la suite que les relations de cause à effet absolues peuvent être considérées comme des relations de cause à effet relatives à un ensemble de variables qui suffit à décrire la réalité empirique.

2.3.1.2 Causalité entre variables et causalité entre propriétés

La seconde différence entre les objets de la caractérisation RB et les théories probabilistes de la causalité concerne la nature des *relata* causaux. Tandis que la caractérisation RB vise la causalité *entre variables*, les théories probabilistes sont des analyses de la causalité *entre propriétés*. Ce point a déjà été mentionné dans le premier chapitre du présent travail.[11]

A l'occasion de cette première discussion, il est apparu d'abord qu'à toute valeur d'une variable correspond de manière univoque une propriété – celle d'être tel que la variable prend ladite valeur. Réciproquement, toute propriété A peut être représentée par la variable binaire V_A qui prend la valeur 1 quand A est instanciée et 0 sinon. Dès lors, nous pouvons traiter la seconde différence entre les objets de la caractérisation RB et ceux des théories probabilistes de la causalité de façon similaire à celle dont nous avons traité la première. Plus précisément, nous pouvons décrire l'objet des théories probabilistes de la causalité dans les termes qui permettent de décrire l'objet de la caractérisation RB. De même qu'une relation de cause à effet absolue peut être considérée comme une relation de cause à effet relative à un ensemble de variables qui suffit à décrire la réalité empirique, de même une relation de cause à effet entre deux propriétés A et B peut être considérée une relation de cause à effet entre les valeurs 1 de deux variables V_A et V_B.

Il reste que le critère de causalité que les réseaux bayésiens causaux véhiculent constitue une caractérisation de la causalité entre variables, et non de la causalité entre valeurs de variables.[12] Toutefois, entre la causalité entre variables et la causalité entre valeurs de variables, il existe un rapport, que nous avons déjà pointé dans le premier chapitre : une variable X cause une variable Y si et seulement s'il existe une valeur x de X et une valeur y de Y telles que la propriété qui correspond à x cause la propriété qui correspond à y. Il en découle que l'énoncé « X cause Y » est équivalent à la disjonction

[11] Voir pp.38 et suivantes.

[12] Voir, à nouveau, la proposition 2.1.

des énoncés de la forme « La propriété qui correspond à la valeur x_i de X cause la propriété qui correspond à la valeur y_j de Y », où x_i et y_j sont des valeurs possibles de X et de Y. Sous l'hypothèse selon laquelle la causalité absolue peut être considérée comme un type particulier de causalité relative, cette équivalence nous permet de préciser et de justifier l'affirmation selon laquelle la notion de causalité qui est à l'oeuvre dans les réseaux bayésiens causaux est grossière. On dira en effet que cette notion est *plus* grossière *que* celle que visent les théories probabilistes de la causalité, et nous le justifierons par ceci qu'un énoncé relatif à celle-là équivaut à une disjonction d'énoncés relatifs à celle-ci.

De ce que nous avons mis en évidence dans les deux derniers paragraphes, il découle que l'énoncé « V_A cause V_B » est équivalent à la disjonction des énoncés de la forme « La propriété qui correspond à v_A cause la propriété qui correspond à v_B ». Autrement dit, nous avons fait apparaître que « V_A cause V_B » est équivalent à la disjonction « A cause B ou A cause non-B ou non-A cause B ou non-A cause non-B ». Sous l'hypothèse selon laquelle V_A cause V_B si et seulement si V_A cause V_B relativement à un ensemble de variables qui suffit à décrire la réalité empirique, l'équivalence entre « V_A cause V_B » et « A cause B ou A cause non-B ou non-A cause B ou non-A cause non-B » a pour premier terme un objet pour la caractérisation RB, et pour second terme un objet d'analyse pour les théories probabilistes de la causalité. A la lumière de cette équivalence, nous pouvons comparer les deux types d'analyses que nous avons présentés dans les sections 2.1 et 2.2.

2.3.2 Comparaison des analyses

En accord avec les analyses que nous venons juste de mener, une comparaison entre la caractérisation RB et les théories probabilistes de la causalité peut prendre la forme d'une comparaison entre les analyses respectives de la disjonction « A cause B ou A cause non-B ou non-A cause B ou non-A cause non-B ». Dans le cadre des méthodes d'inférence aux causes fondées sur les réseaux bayésiens, cette analyse se présente comme la caractérisation de « V_A cause V_B » relativement à un ensemble de variables qui suffit à décrire la réalité empirique. Dans le cours de la comparaison que nous nous apprêtons à mener, nous adopterons deux commodités de langage. D'une part, nous abrégerons la disjonction « A cause B ou A cause non-B ou non-A cause B ou non-A cause non-B », en « (non-)A cause (non-)B ». D'autre part, nous ne préciserons plus que la causalité entre variables est toujours relative à un ensemble de variables qui suffit à décrire la réalité empirique, mais nous le sous-entendrons.

2.3.2.1 Méthodologie

Les analyses de « (non-)A cause (non-)B » par les théories présentées dans la sous-section 2.2 et la caractérisation RB de « V_A cause V_B » peuvent être comparées selon plusieurs critères. Parmi ceux-ci, on distingue en particulier la forme de l'analyse[13], la stratégie utilisée pour tenir compte de tel ou tel type de contre-exemples à l'idée séminale, ou plus simplement ceux de ces contre-exemples que l'analyse prend en charge.

Nous retenons le troisième critère, pour trois raisons non indépendantes. D'abord, les contre-exemples pour une analyse de « (non-)A cause (non-)B » sont les mêmes, moyennant la disjonction, que les contre-exemples pour l'analyse correspondante de « A cause B ». Or, la question des contre-exemples aux analyses proposées pour « A cause B » a gouverné le développement réel des théories probabilistes de la causalité et, surtout, a guidé notre présentation de ces théories. Dès lors, il apparaît naturel de comparer les théories probabilistes et la caractérisation RB sur le critère des contre-exemples. Ensuite, l'étendue du domaine au sein duquel elle est correcte – et donc, ici, les types de contre-exemples qu'elle prend en charge – est ce qui permet d'évaluer une analyse, et ce qu'on en retient *in fine*. Enfin, c'est seulement en procédant à une comparaison selon le troisième critère que nous pourrons tenter de répondre à la question, que nous rappelions au début du chapitre, de la possibilité même d'inférer des relations causales à partir de données probabilistes en utilisant les réseaux bayésiens.

En choisissant ce critère pour la comparaison de la caractérisation RB et des théories probabilistes de la causalité, nous choisissons aussi une façon d'organiser la comparaison. Plus précisément, le troisième critère ayant été choisi, il est naturel de reprendre chacun à son tour les types de contre-exemples que nous avons discutés dans la section 2.2 de présentation des théories probabilistes de la causalité. Pour chacun de ces types, on déterminera si la caractérisation RB prend en charge les contre-exemples de ce type. Pour une plus grande clarté de l'analyse, nous ne reprenons pas l'ordre dans lequel les types de contre-exemples ont été introduits dans la section 2.2.

2.3.2.2 Indépendances trompeuses

En vue de situer efficacement la caractérisation RB dans le portrait de la famille des théories probabilistes de la causalité que nous avons brossé, il est utile de revenir d'abord sur le problème des indépendances trompeuses.

[13]Selon ce critère, il semble que la caractérisation RB est plus proche des théories postérieures à celle de Suppes, que de la théorie de Suppes.

Pour cela, reprenons la conséquence de la caractérisation RB de la causalité que nous avons mise en évidence dans la sous-section 2.1.3 (théorème 2.2). Cette conséquence est la suivante : une condition nécessaire pour qu'une variable X cause une variable Y est que X et Y soient dépendantes en probabilité. En particulier, une variable V_A cause une variable V_B seulement si V_A est dépendante de V_B. Or, sous la signification que nous avons donnée plus haut à la notation V_P pour une propriété P, on a le résultat suivant :

Théorème 2.3 (Dépendance probabiliste entre variables et entre propriétés) *Soit A et B deux propriétés.*
V_A est dépendante de V_B si et seulement si A est dépendante de B.

Ce résultat s'établit de la façon suivante :

Preuve : Soit A et B deux propriétés.
Si A est dépendante de B, alors la valeur 1 de V_A est dépendante de la valeur 1 de V_B et, par la définition 1.15 de l'indépendance probabiliste entre variables, V_A est dépendante de V_B.
Supposons maintenant que V_A est dépendante de V_B. Quatre cas, non exclusifs, sont possibles :

1. la valeur 1 de V_A est dépendante de la valeur 1 de V_B. Dans ce cas, on a immédiatement la conclusion recherchée, à savoir : A est dépendante de B ;
2. la valeur 1 de V_A est dépendante de la valeur 0 de V_B.
   On a alors $p(A|non-B) \neq p(A)$.
   Par le théorème des probabilités totales, on a :
   $p(A) = p(A|non-B).p(non-B) + p(A|B).p(B)$.
   L'inégalité $p(A|non-B) \neq p(A)$ implique alors : $p(A) \neq p(A).p(non-B) + p(A|B).p(B)$, soit :
   $p(A) \neq p(A).p(non-B) + p(B|A).p(A)$, et enfin :
   $p(A) \neq p(A)[p(non-B) + p(B|A)]$.
   En divisant par $p(A)$[14], il vient :
   $1 \neq p(non-B) + p(B|A)$, soit :
   $p(B) \neq p(B|A)$, qui une expression de la dépendance de A et B ;
3. la valeur 0 de V_A est dépendante de la valeur 1 de V_B.
   On a alors $p(non-A|B) \neq p(non-A)$, et la dépendance de A et B découle immédiatement de ce que $p(non-A|B) = 1 - p(A|B)$ et $p(non-A) = 1 - p(A)$;
4. la valeur 0 de V_A est dépendante de la valeur 0 de V_B.
   On a alors $p(non-A|non-B) \neq p(non-A)$. Pour des raisons similaires à celles qui ont été évoquées pour le cas 3., il en découle que

[14]Rappelons que nous ne considérons que des propriétés de probabilité non nulle, en particulier parce qu'il n'y aurait pas de sens clair à parler d'une cause ou d'un effet de probabilité nulle.

$p(A|non - B) \neq p(A)$. Or, on a montré pour le cas 2. que cette inégalité implique la dépendance de A et de B.

Nous avons ainsi établi que V_A est dépendante de V_B si et seulement si A est dépendante de B**.**

De l'équivalence que nous venons d'établir, nous déduisons que, sous la caractérisation RB, l'énoncé « (non-)A cause (non-)B » n'est vrai que si A et B sont dépendantes en probabilité. Autrement dit, dans le cadre de l'inférence aux causes fondées sur les réseaux bayésiens, il n'y a pas de causalité sans dépendance probabiliste. La caractérisation RB échoue donc à prendre en compte la possibilité qu'existent des indépendances trompeuses, de quelque type qu'elles soient plus précisément. De façon plus imagée, la conséquence énoncée dans le théorème 2.2 est à la caractérisation RB ce que sa clause 2. est à la théorie de Suppes.

Tournons-nous, maintenant, du côté des théories probabilistes de la causalité postérieures à Suppes (1970). L'équivalence que nous venons de montrer suffit à établir que, *dans le cas non interactif*, la caractérisation RB de « (non-)A cause (non-)B » est strictement moins satisfaisante que l'analyse de cet énoncé par n'importe laquelle d'entre ces théories. En effet, nous avons vu que les seuls contre-exemples n'impliquant pas de cause interactive sur lesquels certaines de ces théories achoppent sont des cas d'indépendances trompeuses. D'un autre côté, nous avons montré que les théories développées dans Cartwright (1979) et Skyrms (1980) ne sont pas satisfaisantes dans le cas interactif – et nous avons indiqué que seule la théorie proposée dans Cartwright (1989) s'attaque au problème des fourches interactives. Dans ces conditions, nous en venons maintenant aux contre-exemples interactifs à l'idée séminale. Si la caractérisation RB ne les prend pas en charge, il s'ensuivra qu'elle est strictement moins satisfaisante que toutes les théories probabilistes de la causalité qui sont postérieures à Suppes (1970).

2.3.2.3 Corrélations trompeuses avec causes interactives

Conformément à ce que nous avons annoncé plus haut, nous nous en tenons ici aux corrélations trompeuses entre des effets communs à une cause interactive, c'est-à-dire qui ne fait pas écran entre eux. La raison en est que les situations dans lesquelles on trouve de telles corrélations sont les plus simples parmi celles dont c'est le caractère interactif fait problème, en même temps que des situations dont nous avons vu qu'elles peinent déjà à recevoir un traitement correct.

Maintenant, ces corrélations trompeuses sont-elles correctement traitées

par la caractérisation RB ? En vue de répondre à cette question, commençons par remarquer que le résultat 2.3 peut être généralisé au cas de dépendances absolues. Pour le dire explicitement, on a le résultat suivant :

Théorème 2.4 (Généralisation de 2.3) *Soit A, B deux propriétés et* $\boldsymbol{C}$ *un ensemble de propriétés.*
V_A est dépendante de V_B relativement à l'ensemble $\boldsymbol{V_C}$ *des variables qui correspondent aux propriétés de* $\boldsymbol{C}$ *si et seulement si A est dépendante de B relativement à* $\boldsymbol{C}$*.*

Ce résultat admet une preuve similaire à celle que nous avons donnée pour le théorème 2.3.

Considérons maintenant une cause interactive C qui implique une corrélation trompeuse entre deux de ses effets A et B et ne fait pas écran entre eux. Dans ce cas, la proposition « (non-)A cause (non-)B » est fausse. La question est de savoir si la caractérisation RB conduit à cette conclusion ou, de façon équivalente, si elle conduit à conclure que V_A ne cause pas V_B. Pour répondre à cette question, rappelons que, sous la caractérisation RB, V_A cause V_B si et seulement si les deux variables sont dépendantes relativement à l'ensemble des causes directes de V_A. Dans le cas qui nous occupe, et par hypothèse sur la situation, V_A a une seule cause directe : V_C. Dans ces conditions, la question est de savoir si V_A est dépendante de V_B relativement à V_C. Par le résultat 2.4, V_A est dépendante de V_B relativement à V_C si et seulement si A est dépendante de B relativement à C. Or, par hypothèse sur la situation, A est bien dépendante de B relativement à C. Donc la caractérisation RB conclut à tort que V_A cause V_B, c'est-à-dire que « (non-)A cause (non-)B » est faux.

Il apparaît donc que la caractérisation RB ne traite pas correctement le problème des corrélations trompeuses entre effets d'une même cause qui ne fait pas écran entre eux. Conformément à ce que nous avons annoncé, il en découle que la caractérisation RB se heurte à plus de contre-exemples que n'importe laquelle des théories probabilistes de la causalité développées après 1970. Dès lors, et toujours en vue de la situer dans le champ des théories probabilistes de la causalité, c'est avec la théorie proposée par Suppes qu'il convient de comparer maintenant la caractérisation RB. Pour ce qui est de cette comparaison, nous commençons par déterminer si la caractérisation RB traite correctement au moins autant de cas que la théorie 2.2 de Suppes. En d'autres termes, nous déterminons si la caractérisation RB permet d'identifier comme trompeuses les corrélations trompeuses identifiées comme telles par la théorie de Suppes. Ces corrélations trompeuses sont précisément des corrélations entre effets communs à une cause qui n'est pas interactive.

2.3.2.4 Corrélations trompeuses entre effets d'une même cause qui n'est pas interactive

Dans ce paragraphe, nous envisageons une corrélation trompeuse entre deux effets A et B d'une propriété C qui fait écran entre eux. A titre de rappel, A et B sont par exemple les propriétés de développer un cancer du poumon et d'avoir les doigts jaunis, et C la propriété d'être fumeur. Dans une telle situation, l'énoncé « (non-)A cause (non-)B » est faux, et la question est de savoir si la caractérisation RB conduit bien à cette conclusion. En termes de variables, la question est de savoir si la caractérisation RB conduit bien à la conclusion selon laquelle V_A ne cause pas V_B.

Sous la caractérisation RB (proposition 2.1), V_A cause V_B si et seulement si V_A et V_B sont dépendantes relativement à l'ensemble des causes directes de V_A. Or, parmi ces causes directes figure soit V_C, soit un effet I de V_C. Dans le premier cas, l'hypothèse selon laquelle C fait écran à la dépendance entre A et B implique que V_A et V_B sont indépendantes relativement à l'ensemble des causes directes de V_A. Dans le second cas, il faut utiliser le théorème des probabilités totales pour aboutir à la même conclusion :

Preuve. Soit $\mathbf{CD}_A$ l'ensemble des causes directes de V_A différentes de I. Soit aussi v_A une valeur de V_A, i une valeur de I, v_B une valeur de V_B et $\mathbf{cd}_A$ une valeur de $\mathbf{CD}_A$.
Par le théorème des probabilités totales on a :

$$\begin{aligned} p(v_A|v_B.i.\mathbf{cd}_A) = \quad & p(v_A|v_B.i.\mathbf{cd}_A.(V_C = 1)).p(V_C = 1)+ \\ & p(v_A|v_B.i.\mathbf{cd}_A.(V_C = 0)).p(V_C = 0). \end{aligned}$$

Par hypothèse, A et B sont indépendantes relativement à C. Il en découle par le théorème 2.4 que V_A et V_B sont indépendantes relativement à V_C.
En conséquence, pour toute valeur v_C de V_C, on a :
$p(v_A|v_B.i.\mathbf{cd}_A.v_C) = p(v_A|i.\mathbf{cd}_A.v_C)$.
Du coup, on a :

$$\begin{aligned} p(v_A|v_B.i.\mathbf{cd}_A) = \quad & p(v_A|i.\mathbf{cd}_A.(V_C = 1)).p(V_C = 1)+ \\ & p(v_A|i.\mathbf{cd}_A.(V_C = 0)).p(V_C = 0). \end{aligned}$$

En utilisant à nouveau le théorème des probabilités totales, il vient :
$p(v_A|v_B.i.\mathbf{cd}_A) = p(v_A|i.\mathbf{cd}_A)$.
En conséquence, V_A et V_B sont indépendantes relativement à l'ensemble des causes directes de V_A**.**

Ainsi apparaît-il que V_A est indépendante de V_B relativement à l'ensemble des causes directes de V_A dans tous les cas – c'est-à-dire que V_C soit ou non l'une de ces causes *directes*. En conséquence, V_A ne cause pas V_B selon la caractérisation RB de la causalité. De façon plus générale, il apparaît que la caractérisation RB est adéquate pour le type de cas que nous envisageons dans le présent paragraphe.

Nous venons de montrer que la caractérisation RB traite correctement les cas de corrélations trompeuses entre effets d'une même cause qui fait écran entre eux. Pour le dire autrement, elle prend en charge tous ceux des contre-exemples à l'idée séminale que la théorie de Suppes traite correctement. Reste donc à déterminer si la caractérisation RB prend en charge des contre-exemples sur lesquels la théorie de Suppes achoppe – ou si au contraire la caractérisation RB se heurte exactement aux difficultés que nous avons mises plus haut (dans la sous-section 2.2.3) pour la théorie de Suppes. Celles de ces difficultés qui n'impliquent pas de cause interactive correspondent aux corrélations trompeuses entre effets de plusieurs causes d'abord, aux corrélations entre effets et causes ensuite, et au paradoxe de Simpson enfin. Nous abordons les trois points dans cet ordre.

2.3.2.5 Corrélations trompeuses entre effets de plusieurs causes

Commençons, donc, par les corrélations trompeuses entre effets de plusieurs causes. Nous avons vu plus haut (dans le paragraphe 2.2.3.3) que ces corrélations sont de deux sous-types : d'une part les corrélations entre des effets qui ont *plusieurs* causes en commun, d'autre part les corrélations entre deux variables causalement indépendantes et dont l'une a plusieurs causes directes dont aucune ne suffit à déterminer la valeur qu'elle prend.

La distinction entre ces deux sous-types n'est pas importante ici. En effet, nous avons mis en évidence plus haut (toutjours dans le paragraphe 2.2.3.3) que la difficulté rencontrée ici par la théorie de Suppes découle en fait de ce qu'elle n'identifie une corrélation entre deux propriétés comme trompeuse que si *une* troisième propriété suffit à faire écran entre les deux premières. Or il nous semble clair que la difficulté ainsi décrite est surmontée par la caractérisation RB. Sous cette caractérisation, en effet, une relation de cause à effet est une relation de dépendance probabiliste qui résiste à la conditionalisation sur l'ensemble de *toutes* les causes directes de la cause. Dans ces conditions, il apparaît que la caractérisation RB traite correctement de ce premier type de cas non interactifs qui posent problème pour la théorie de Suppes.

2.3.2.6 Corrélations trompeuses entre effets et causes

De même que les corrélations trompeuses entre effets de plusieurs causes, les corrélations trompeuses entre les effets et leurs causes posent problème dans le cadre de la théorie de Suppes. Plus précisément, nous avons vu (dans le paragraphe 2.2.3.2) que la théorie de Suppes pose deux problèmes sur ce point. En premier lieu, elle recourt à une solution temporelle dont le sens

n'est pas complètement clair relativement à la causalité *générique*. En second lieu, elle implique que la causalité générique est asymétrique, alors que nous avons vu qu'elle fait, parfois, des cycles.

A l'instar des théories probabilistes postérieures à Suppes (1970), la caractérisation RB ne se heurte à aucune de ces deux difficultés. D'une part, elle ne comporte pas une analyse temporelle du sens des relations de cause à effet. D'autre part, elle n'implique pas que la causalité générique est une relation asymétrique. En effet, il est possible que deux variables soient dépendantes relativement à l'ensemble des causes directes de l'une *et* relativement à l'ensemble des causes directes de l'autre. On notera que ce dernier point est compatible avec l'acyclicité des graphes bayésiens : une chose est la caractérisation de la causalité sur laquelle reposent les méthodes d'inférence aux causes qui utilisent les réseaux bayésiens, autre chose est le domaine dans lequel ces méthodes recherchent leurs résultats. Seule la première chose nous intéresse dans le présent chapitre, et elle est telle que la causalité générique n'est pas supposée asymétrique. La caractérisation RB ne présente donc pas les deux défauts de la théorie de Suppes relativement aux corrélations entre effets et causes.

D'un autre côté, la caractérisation RB n'implique pas que la relation de causalité est symétrique. En effet, deux variables peuvent être dépendantes relativement à l'ensemble des causes directes de l'une sans être dépendantes relativement à l'ensemble des causes directes de l'autre. Plus généralement, la caractérisation RB n'implique rien relativement à la réciprocité ou à la non réciprocité des relations de cause à effet. En conséquence, nous concluons qu'elle traite correctement les corrélations trompeuses entre effets et causes.

2.3.2.7 Paradoxe de Simpson

Il nous reste à achever notre comparaison de la caractérisation RB avec les théories probabilistes de la causalité en général et avec celle de Suppes en particulier. Pour cela, il nous faut déterminer si la première traite correctement les corrélations trompeuses impliquées par le paradoxe de Simpson, c'est-à-dire les corrélations dont le sens est trompeur.

En vue de déterminer si la caractérisation RB traite correctement les indépendances trompeuses relevant du paradoxe de Simpson, il faut revenir à ceci que ce qui est comparable est l'analyse de « (non-)A cause (non-)B » par les théories probabilistes de la causalité et la caractérisation RB de « V_A cause V_B ». Par ailleurs, rappelons aussi que le paradoxe de Simpson soulève en fait le problème de l'identification de celle d'une propriété ou de sa négation qui doit compter comme cause d'une troisième propriété. Le paradoxe de Simpson, et avec lui les corrélations trompeuses qu'il implique,

se définissent en termes de propriétés. Dans ces conditions, il ne peut pas être défini dans le cadre d'une notion RB de causalité. Plus précisément, l'objet de la caractérisation RB de la causalité est trop grossier pour que la caractérisation soit sensible aux distinctions engagées par le paradoxe de Simpson. Le paradoxe de Simpson ne peut donc pas être un problème pour la caractérisation RB de la causalité.

Ainsi, non seulement les contre-exemples à l'idée séminale pris en charge par la théorie de Suppes le sont également par la caractérisation RB, mais encore celle-ci traite correctement certains des contre-exemples sur lesquels celle-là achoppe : les corrélations trompeuses entre effets de plusieurs causes d'une part, et d'autre part les corrélations trompeuses entre effets et causes. De l'autre côté il est apparu que la caractérisation RB achoppe sur strictement plus de contre-exemples que n'importe laquelle des théories probabilistes postérieures à Suppes (1970). Dans ces conditions, la caractérisation RB de l'énoncé « V_A cause V_B » – qui, rappelons-le, doit être compris comme relatif à un ensemble de variables qui suffit à décrire la réalité physique – s'inscrit juste après Suppes (1970) dans l'ordre historique des analyses de « (non-)A cause (non-)B » par les théories probabilistes de la causalité.

2.3.3 Conséquences pour l'inférence aux causes

La comparaison avec les différentes théories probabilistes de la causalité a rendus saillants certains traits de la caractérisation RB de la causalité. Pour en finir avec l'analyse de cette caractérisation, nous montrons comment certains de ces traits ont des conséquences sur l'inférence aux causes fondées sur les réseaux bayésiens. En particulier, nous tentons de répondre à la question initiale de la possibilité même d'inférer, grâce aux réseaux bayésiens, des relations de cause à effet à partir de données probabilistes d'observation.

Notre analyse des conséquences de ce que nous avons dit de la caractérisation RB de la causalité comporte deux temps. Ces deux temps correspondent à deux des traits de la caractérisation dont dépend la possibilité d'inférer des causes grâce aux réseaux bayésiens. Le premier de ces traits est le suivant : la caractérisation RB de la causalité ne prend pas en charge les indépendances trompeuses. Pour le dire autrement, quand la condition de Markov causale et l'hypothèse de fidélité sont satisfaites, la dépendance probabiliste est une condition nécessaire de causalité. Ce trait se présente comme un bon candidat pour l'explication de la possibilité même d'inférer des relations de cause à effet à partir de données probabilistes. On se souvient en effet que c'est précisément pour prendre en compte les indépendances trompeuses que des théories probabilistes circulaires ont été introduites.

Avant d'en venir aux conséquences du théorème 2.2 pour l'inférence aux causes, il convient de rappeler que les méthodes d'inférence aux causes fondées sur les réseaux bayésiens visent les relations de cause à effet *directes*. Nous ne discutons pas précisément ici la différence que cela fait par rapport à la caractérisation que nous avons analysée plus haut dans le chapitre. Plutôt, nous remarquons que la causalité est une condition nécessaire de causalité directe, d'où il découle que la dépendance probabiliste est une condition nécessaire de causalité directe. Or, ce fait, qui dépend étroitement de la caractérisation RB de la causalité, est fondamental dans les algorithmes d'inférence aux causes qui nous intéressent ici. En particulier, l'algorithme PC de Spirtes, Glymour et Scheines commence par identifier tous les couples de variables indépendantes[15], et conclure que les deux variables qui composent un tel couple ne sont pas en relation de cause à effet.

Cela, toutefois, ne serait rien sans le résultat plus général selon lequel, sous la condition de Markov causale et l'hypothèse de fidélité, deux variables X et Y sont en relation de cause à effet dans l'ensemble $\mathbf{V}$[16] si et seulement si elles sont dépendantes relativement à tous les sous-ensembles de $\mathbf{V}\backslash\{X, Y\}$.[17] On pourrait reprendre la démonstration de ce résultat[18] et en produire une analyse qui fasse apparaître comment le résultat lui-même s'articule avec ce que nous avons dit plus haut de la caractérisation RB de « X cause Y ». Nous ne le faisons pas ici, pour deux raisons. La première est que les conclusions qu'on peut espérer tirer de cette enquête ne sont pas la hauteur de l'investissement qu'elle représente. En effet, il nous semble déjà relativement clair que, parmi les traits de la caractérisation RB que nous avons mis en évidence plus haut, c'est avec l'absence de prise en compte des indépendances trompeuses que le résultat auquel nous faisons allusion ici a le plus à voir. Conformément à cette analyse, nous considérons que la conséquence de la caractérisation RB qui est énoncée dans le théorème 2.2 est essentielle à la possibilité d'inférer des causes à partir de données probabilistes grâce aux réseaux bayésiens.

La seconde raison pour laquelle nous n'analysons pas la preuve du résultat que nous venons de mentionner est la suivante : ce résultat lui-même ne serait d'aucune utilité pour inférer des relations causales à partir de données probabilistes si la caractérisation RB de la causalité ne possédait pas un autre des traits qui sont apparus plus haut. Plus explicitement, ce résultat ne serait rien si la notion de causalité visée par la caractérisation RB, et donc par les

[15] Voir la clause B.) pour $n = 0$; Spirtes et al. (1993) pp. 84–85.

[16] Par là nous entendons que ls deux variables sont adjacentes dans le graphe causal sur $\mathbf{V}$, c'est-à-dire que A cause directement B dans $\mathbf{V}$ ou B cause directement A dans $\mathbf{V}$.

[17] Spirtes et al. (1993), Théorème 3.4 p. 82.

[18] Elle est donnée dans Spirtes et al. (1990).

méthodes d'inférence aux causes fondées sur les réseaux bayésiens, n'était pas relative.

En effet, les analyses menées dans les sous-sections 2.1.3 et 2.3.2 font apparaître que, pour la caractérisation RB, une relation de cause à effet est une relation de dépendance probabiliste qui résiste à la conditionalisation. D'après le résultat que nous évoquions dans le dernier paragraphe, il en va de même de la causalité directe. Dès lors, l'inférence aux causes devient la recherche de dépendances probabilistes qui ne disparaissent pas par conditionalisation. Or, comment conclure qu'une dépendance ne disparaît pas par conditionalisation si l'ensemble des conditionnants à envisager est indéterminé et potentiellement infini ? Savoir que c'est la conditionalisation sur les causes directes de X qui permet de déterminer si X cause Y n'avance à rien dès lors que ce sont précisément les relations de cause à effet qu'on cherche à mettre au jour. Dans ces conditions, ce qui compte, et qui rompt la circularité de l'analyse, est la possibilité d'envisager *toutes* les indépendances probabilistes relatives. Cela est permis par ceci que les relations de cause à effet représentées par les graphes bayésiens sont relatives à un ensemble de variables. Nous voyons donc dans ce caractère relatif une seconde condition de possibilité de l'inférence aux causes à partir de données probabilistes quand nos meilleures théories probabilistes de la causalité – et avec elles la caractérisation RB – sont circulaires. Le caractère relatif des relations causales qu'on infère apparaît alors comme le prix à payer pour la possibilité de cette inférence.

2.4 Conclusion

Dans la première section du chapitre qui s'achève, nous avons mis au jour et analysé la caractérisation RB de la causalité – c'est-à-dire la caractérisation qui découle de la condition de Markov causale et de l'hypothèse de fidélité. Nous avons fait apparaître en particulier que, sous cette caractérisation, la dépendance probabiliste est une condition nécessaire de dépendance causale.

Aussi important soit-il, ce résultat ne nous a pas suffi pour mener à bien la comparaison entre la caractérisation RB et les théories probabilistes de la causalité présentées dans la deuxième section. En effet, il est apparu dès le début de la troisième section que les deux types d'analyses diffèrent d'abord par ce sur quoi elles portent. En conséquence, nous avons dû construire un énoncé complexe qu'elles puissent viser ensemble. La comparaison a été menée relativement à l'analyse de cet énoncé complexe. En outre, elle a pris pour critère la capacité à rendre compte correctement des différents types de contre-exemples à l'idée séminale de caractériser la causalité par l'augmen-

tation de probabilité.

La comparaison a fait apparaître que, *mutatis mutandis*, la caractérisation RB traite correctement tous les cas de causalité pris en charge par la théorie de Suppes, ainsi que certains cas non pris en charge par cette théorie. Néanmoins, il est apparu ensuite que, ne tenant pas compte de l'existence d'indépendances trompeuses, la caractérisation RB n'est en mesure d'analyser correctement qu'un ensemble de cas strictement contenu dans celui des cas traités par n'importe laquelle des théories probabilistes de la causalité postérieures à Suppes (1970). Dans l'histoire rationnellement reconstruite des théories probabilistes de la causalité, la caractérisation RB prend place juste après la théorie proposée dans Suppes (1970), mais strictement après elle.

Dans la dernière sous-section du chapitre, nous avons montré que la situation de la caractérisation RB dans le champ des théories probabilistes du concept de cause explique en partie pourquoi les réseaux bayésiens permettent d'inférer des causes à partir de données probabilistes. Pour l'autre partie, nous avons défendu que le caractère relatif – à un ensemble de variables – de la causalité dans un réseau bayésien causal joue également un rôle essentiel.

Selon les lignes que nous venons de rappeler, nous avons répondu à la question des modalités de l'inférence aux causes autorisée par les réseaux bayésiens telle qu'elle se pose du point de vue de l'analyse conceptuelle. Il nous reste donc à discuter le second des points de vue que nous envisagions en introduction au présent chapitre, celui de la méthodologie de l'inférence aux causes fondée sur les réseaux bayésiens. Nous nous y attelons dans le prochain chapitre.

Chapitre 3

Réseaux bayésiens et inférence causale

Dans le chapitre qui commence, nous abordons la question de la contribution des réseaux bayésiens à l'épistémologie de la causalité générique depuis le point de vue de la *méthodologie* de l'inférence causale. Il s'agit non plus de préciser le rapport entre les critères de causalité qui sont à l'oeuvre dans les réseaux bayésiens causaux et les analyses du concept de cause générique, mais de situer l'inférence aux causes fondée sur les réseaux bayésiens dans le champ même des méthodes d'inférence causale. Ce projet est central relativement à la question épistémologique qui nous occupe dans cette première partie de notre travail. Mais nous avons vu dans l'introduction qu'il se motive aussi plus spécifiquement. En effet, l'idée est assez répandue selon laquelle les réseaux bayésiens permettraient d'*induire* des relations de cause à effet, c'est-à-dire d'acquérir des connaissances causales générales à partir de l'observation de faits particuliers. Si tel est bien le cas, les réseaux bayésiens révolutionnent la recherche des causes génériques, dont il a semblé clair au moins depuis Popper (1934) qu'elle doit suivre des voies hypothético-déductives.

De ce qui motive spécifiquement la question de la contribution des réseaux bayésiens à l'inférence causale, il suit deux conséquences pour le chapitre qui commence. Selon la première, ce qui doit nous intéresser au premier chef ici est la *logique* de l'inférence causale, et donc plus généralement ses principes. En vue d'expliquer mieux ce dont il est question ici, il convient de distinguer entre trois choses :

- les principes de l'inférence causale, c'est-à-dire ce en quoi consiste l'inférence, sa structure générale ;
- les procédures d'inférence causale, c'est-à-dire la liste ordonnée des tâches qui sont réalisées afin d'obtenir une conclusion causale à partir des prémisses, ici probabilistes ;

– la mise en oeuvre de ces procédures d'inférence causale, c'est-à-dire les conditions concrètes de la réalisation de ces tâches.

Les questions des procédures et de la mise en oeuvre sont abordées soit au titre de préalables à la mise au jour des principes, soit au titre de compléments à la description des principes.

La seconde conséquence que nous tirons de la nature de notre motivation spécifique à poser la question de la contribution des réseaux bayésiens à la méthodologie de l'inférence causale est la suivante : nous adoptons une méthode d'analyse comparative. Plus explicitement : nous rendons saillantes les caractéristiques propres à l'inférence causale fondée sur les réseaux bayésiens à l'occasion de sa comparaison avec l'inférence causale plus traditionnelle qu'elle vient concurrencer. Ainsi que nous l'avons suggéré déjà, l'inférence aux causes génériques est traditionnellement hypothético-déductive.

Finalement, de même que celle du chapitre 2, l'organisation du chapitre qui commence s'ordonne au projet de comparaison. Dans une première section, nous présentons les méthodes d'inférence causale fondées sur les réseaux bayésiens plus extensivement que nous n'avons eu à la faire jusqu'à présent. Dans une deuxième section, nous présentons les méthodes d'inférence causale plus traditionnelles que nous aurons identifiées comme celles auxquelles les méthodes fondées sur les réseaux bayésiens doivent être comparées. La troisième section est consacrée à mener effectivement la comparaison. A l'issue de ces analyses, la question se posera de la place qu'on peut accorder aux réseaux bayésiens dans les méthodes d'inférence causale ; dans une quatrième section nous envisageons des éléments de réponse à cette question. Une courte dernière section est consacrée à une présentation synthétique de nos résultats.

3.1 Inférence causale fondée sur les réseaux bayésiens

L'inférence causale fondée sur les réseaux bayésiens (« inférence causale RB » dans la suite) existe le plus visiblement sous la forme des algorithmes d'inférence causale que nous avons mentionnés dans la sous-section 1.2.1. En conséquence, nous adoptons dans la section qui commence la stratégie suivante : présenter d'abord les procédures d'inférence causale RB dont ces algorithmes participent, en tirer une description des principes de l'inférence causale RB, compléter la description du domaine de l'inférence causale RB en traitant la question de sa mise en oeuvre.

3.1.1 Procédure d'inférence causale RB

Nous avons choisi de commencer par présenter ici les procédures d'inférence causale parce que les algorithmes d'inférence causale fondés sur les réseaux bayésiens nous donnent la prise la plus ferme sur ce qu'est l'inférence causale RB. Dans ces conditions, il est naturel que nous commencions par présenter précisément ces algorithmes.

3.1.1.1 Algorithmes d'inférence causale RB

Ainsi que nous l'avons indiqué déjà dans la sous-section 1.2.1, les algorithmes d'inférence aux causes fondés sur les réseaux bayésiens (« algorithmes RB » dans la suite) se sont développés en deux séries parallèles : d'un côté, les algorithmes IC et IC* de Verma et Pearl, de l'autre les algorithmes SGS, PC et PC*, et CI et FCI de Spirtes, Glymour et Scheines. Parmi ces algorithmes, nous présentons le seul algorithme PC de Spirtes, Glymour et Scheines. La raison pour laquelle nous pouvons nous contenter de présenter un seul des nombreux algorithmes RB est la suivante : dans tous les cas, le principe de l'inférence causale – qui est ce qui nous intéresse finalement – est le même. Maintenant, le choix de l'algorithme PC repose sur deux considérations non indépendantes. D'une part PC est l'algorithme qu'utilisent les plus connus des programmes informatiques d'inférence causale RB – les programmes de la famille TETRAD. D'autre part PC, et plus généralement les algorithmes de Spirtes, Glymour et Scheines, ont donné lieu aux débats les plus nourris de la littérature philosophique consacrée aux algorithmes RB.[1]

L'algorithme PC est introduit dans Spirtes et al. (1991) et présenté dans Spirtes et al. (1993)[2]. C'est à cette présentation que nous nous référons. Il s'agit d'un algorithme qui vise l'inférence aux causes dans les ensembles de variables « causalement suffisants », c'est-à-dire tels que toute variable qui est une cause commune à au moins deux variables de l'ensemble est elle-même une variable de l'ensemble. Cette restriction est levée par l'algorithme CI[3]. En nous intéressant à PC plutôt qu'à CI, nous limitons les considérations techniques au minimun nécessaire pour traiter les problèmes qui nous intéressent, au niveau d'abstraction depuis lequel nous les abordons. Corrélativement au choix de présenter PC, la question de la sélection des variables qu'il convient de considérer lors de l'étude de la causalité dans un système réel donné n'est pas abordée dans le présent chapitre.

[1] En particulier, la querelle qui se déroule au long de Humphreys et Freedman (1996), Spirtes et al. (1997), Korb et Wallace (1997), Freedman et Humphreys (1999) concerne spécifiquement l'algorithme PC et les programmes de la famille TETRAD.

[2] Spirtes et al. (1993) pp. 84–85.

[3] Spirtes et al. (1993) pp. 139–140.

L'entrée (*input*) de l'algorithme PC est une distribution de probabilités p sur un ensemble de variables $\mathbf{V}$; son résultat (*output*) est un graphe acyclique partiellement orienté – ou « patron » – sur $\mathbf{V}$. PC se compose de quatre instructions. Ces instructions ne nous intéressent pas ici en tant que telles, mais seulement en tant qu'elles composent une procédure dont nous extrairons le principe de l'inférence causale RB. Dès lors, nous renvoyons le lecteur intéressé par leur détail technique à Spirtes et al. (1993)[4] et nous nous contentons d'expliquer ce qu'il s'agit de faire à chaque étape.

Étant donnée une distribution de probabilités sur un ensemble de variables $\mathbf{V}$, PC commande de :
Étape 1 : Former sur $\mathbf{V}$ le graphe non orienté complet (c'est-à-dire tel que chaque variable de $\mathbf{V}$ est reliée à chaque autre variable de $\mathbf{V}$ par une arête), et noter C_1 ce graphe.
Étape 2 : Retirer de C_1 toutes les arêtes entre deux variables X et Y pour lesquelles il existe un ensemble de variables de $\mathbf{V} \setminus \{X, Y\}$ relativement auquel elles sont indépendantes, et noter C_2 le graphe obtenu.
Étape 3 : Remplacer par $X \rightarrow Y \leftarrow Z$ tout sous-graphe $X - Y - Z$ de C_2, tel que X et Z sont indépendants relativement à Y, et noter C_3 le graphe obtenu.
Étape 4 : Orienter toutes les arêtes de C_3 qui peuvent l'être sans qu'aucun sous-graphe de la forme $X \rightarrow Y \leftarrow Z$, ni aucun cycle ne soit créé, et noter $GI_{\mathbf{V}}$ le résultat.
$GI_{\mathbf{V}}$ est le résultat donné par PC.

3.1.1.2 Identifier les indépendances probabilistes relatives

Avant d'aller plus loin, il convient de remarquer que la procédure d'inférence causale que nous venons de décrire requiert que soient identifiées les indépendances probabilistes relatives entre les variables de l'ensemble $\mathbf{V}$ considéré. Pour des raisons computationnelles évidentes, ces indépendances probabilistes ne sont pas identifiées toutes avant le commencement de la procédure PC. Elles sont identifiées à l'occasion de l'étape 2 et dans l'ordre suivant : d'abord les indépendances probabilistes absolues, ensuite les indépendances probabilistes relatives à *une* variable entre variables qui ne sont pas indépendantes absolument, puis les indépendances probabilistes relatives à *deux* variables entre variables qui ne sont indépendantes ni absolument ni relativement à *une* variable...

Pour ce qui est de la procédure d'identification des indépendances probabilistes relatives, il convient de distinguer deux cas. Si on connaît la dis-

[4] Spirtes et al. (1993) p. 84–85.

tribution de probabilités sur l'ensemble de variables et dans la population considérés, alors identifier les indépendances probabilistes revient à comparer les valeurs de probabilités conditionnelles. Toutefois, dans la pratique réelle de l'inférence causale, il arrive rarement qu'on connaisse la distribution de probabilités dans la population. Positivement, on dispose en général de fréquences relatives dans un échantillon de la population. Il faut alors recourir à des tests statistiques d'hypothèses de la forme « X est indépendant de Y relativement à **V'** *dans la population considérée* » où X et Y sont des variables de **V** et **V'** un sous-ensemble de $\mathbf{V} \setminus \{X, Y\}$. Ces tests dépendent de la forme supposée de la distribution des différentes propriétés dans la population. On notera que l'indépendance probabiliste relative est équivalente à l'annulation des corrélations partielles sous les deux hypothèses de normalité des distributions des variables et de linéarité des relations fonctionnelles entre les variables.

3.1.2 Principes de l'inférence causale RB

Maintenant que nous avons présenté une procédure d'inférence causale RB, nous pouvons en venir à ce qui nous intéresse le plus directement, à savoir les principes qui guident l'inférence. En d'autres termes, nous allons pouvoir mettre au jour ce en quoi consiste l'inférence causale RB et, au-delà, en déterminer la nature. En vue de cela, nous commençons par revenir sur les étapes 1 à 4 de l'algorithme PC et par montrer comment elles s'articulent aux hypothèses corrélatives de la notion de réseau bayésien causal (qui ont été mises au jour dans le chapitre 1).

3.1.2.1 Algorithme PC et hypothèses corrélatives des réseaux bayésiens causaux

Une fois formé le graphe non orienté complet sur l'ensemble de variables considéré (étape 1), l'algorithme PC procède en deux temps. Dans un premier temps (étape 2), l'arête entre deux variables est retirée si ces variables sont indépendantes en probabilité, ou si leur dépendance disparaît par conditionalisation sur un ensemble de variables de **V** auquel elles n'appartiennent pas. Ainsi, dans ce premier temps, on considère que deux variables sont en relation de cause à effet si et seulement si 1) elles sont dépendantes en probabilités et 2) il n'existe aucun ensemble de variables qui fait écran entre elles. Nous retrouvons ici l'aspect de la caractérisation RB de la causalité que nous avons mis en évidence dans la sous-section 2.1.3 et sa généralisation mentionnée dans la sous-section 2.3.3. De manière générale, ce premier moment vise à ne retenir comme causales que les relations qui le sont sous la condition

de Markov causale et l'hypothèse de fidélité.

Dans un second temps (étapes 3 et 4), certaines des arêtes du graphe (non orienté) obtenu sont orientées. Elles le sont selon deux principes, qui correspondent :

- pour le premier : au fait que, sous la condition de Markov causale et l'hypothèse de fidélité, un sous-graphe X — Y — Z a l'orientation $X \rightarrow Y \leftarrow Z$ si et seulement si X et Z sont indépendants relativement à Y ;
- pour le second : à l'hypothèse d'acyclicité des graphes bayésiens.

Il apparaît alors que l'inférence causale RB consiste à construire le patron $GI_{\mathbf{V}}$ qui représente toute l'information causale relative à $\mathbf{V}$ qu'on peut tirer de la distribution de probabilités initiale sous l'hypothèse d'acyclicité, la condition de Markov causale et l'hypothèse de fidélité. Ainsi, $GI_{\mathbf{V}}$ est composé de toutes et seulement les arêtes, orientées ou non, que partagent tous les graphes orientés acycliques qui représentent la distribution de probabilités p sur $\mathbf{V}$. Les flèches orientées de $GI_{\mathbf{V}}$ sont des flèches qui figurent dans tous les graphes orientés acycliques représentant p. Les arêtes non orientées de $GI_{\mathbf{V}}$ reçoivent dans ces différents graphes des orientations différentes – mais toujours dans la limite qui consiste à ne créer ni sous-graphe de la forme $X \rightarrow Y \leftarrow Z$, ni cycle. Cela mis au jour, nous pouvons nous tourner vers la question de la nature de l'inférence causale RB.

3.1.2.2 Nature de l'inférence causale RB

L'inférence causale qu'on mène en utilisant les réseaux bayésiens est-elle inductive ? En première approche, la réponse à cette question est positive. D'une part nous avons commencé ce chapitre en mentionnant qu'elle était souvent considérée comme telle. D'autre part et surtout la description des procédures d'inférence causale RB fait apparaître que la connaissance causale acquise en utilisant les réseaux bayésiens est tirée des seules données d'observation, indépendamment de toute théorie.

Cette première analyse, toutefois, demande à être affinée. Pour le comprendre, on peut s'arrêter un instant sur les raisons que nous avons de considérer que les inférences RB sont inductives. Ces raisons consistent exactement dans ceci que la connaissance des causes génériques est tirée seulement de données issues de l'observation de cas particuliers. Il semble donc que ce qui autorise à qualifier d'« inductives » les inférences RB est qu'elles tirent des conclusions générales de prémisses particulières et qu'elles les tirent directement – c'est-à-dire en particulier indépendamment de toute théorie. L'inductivité est alors *a-théoricité* ; elle se situe au plan du rapport entre la nature des prémisses et celle de la conclusion.

Mais il existe un autre plan auquel le qualificatif d'« inductif » fait sens : celui du rapport logique entre les prémisses du raisonnement et sa conclusion.[5] Plus explicitement, ce plan est celui du rapport logique entre un ensemble d'indépendances probabilistes $\mathbf{I_V}$ que l'algorithme PC prend pour entrée et le patron causal $GI_{\mathbf{V}}$ qu'il donne comme résultat.

Nous venons de voir (dans le paragraphe 3.1.2.1) qu'une inférence RB consiste à construire le graphe qui représente toute l'information causale qu'on peut tirer des données traitées sous l'hypothèse d'acyclicité, la condition de Markov causale et l'hypothèse de fidélité. Ainsi, $GI_{\mathbf{V}}$ représente toutes et exactement les relations de cause à effet directes entre les variables de $\mathbf{V}$ qui sont impliquées par les indépendances probabilistes relatives qui appartiennent à $\mathbf{I_V}$ sous l'hypothèse d'acyclicité, la condition de Markov causale et l'hypothèse de fidélité.

Il apparaît alors que, sous les trois hypothèses corrélatives des réseaux bayésiens causaux, le rapport entre $\mathbf{I_V}$ et $GI_{\mathbf{V}}$ est de *nécessité* : si les indépendances probabilistes relatives entre les variables de $\mathbf{V}$ sont celles qui appartiennent à $\mathbf{I_V}$, alors *nécessairement* les relations de cause à effet directes qui sont représentées par $GI_{\mathbf{V}}$ existent. Pour être plus concis, on dira que l'information causale véhiculée par le résultat des algorihtmes d'inférence causale est nécessairement vraie si l'est l'entrée probabiliste de l'algorithme. Dans la mesure où leur conclusion découle nécessairement de l'ensemble traité des données probabilistes, il faut reconnaître que les inférences causales RB sont déductives.

Nous avons expliqué pourquoi les inférences causales RB sont inductives, puis montré en quel sens elles sont déductives. Les deux qualificatifs ne sont pas ici incompatibles puisqu'ils ne portent pas sur la même chose. Le premier vise le rapport entre la nature des prémisses et la nature de la conclusion d'une inférence causale RB ; le second vise le rapport logique entre l'entrée et le résultat d'un algorithme d'inférence causale RB. Il convient donc de distinguer soigneusement les objets des qualificatifs « inductif » et « déductif » en tant qu'ils s'appliquent à l'inférence causale RB. Nous le faisons grâce à la distiction entre stratégie et nature de l'inférence : nous dirons que la stratégie de l'inférence causale RB est inductive là où l'inférence elle-même est déductive. Le point des principes de l'inférence causale RB étant ainsi traité, il convient que nous nous arrêtions un moment sur la question de leur mise en oeuvre.

[5] L'idée selon laquelle l'induction fait l'objet de plusieurs caractérisations classiques qui ne coïncident pas toujours est suggérée par Vickers (2006), section 1 en particulier.

3.1.3 Mise en oeuvre : les programmes TETRAD

La stratégie de l'inférence causale RB est inductive et les procédures d'inférence s'organisent autour d'algorithmes. Dans ces conditions, l'inférence causale RB se prête particulièrement bien à l'automatisation. En fait, elle en est même inséparable : l'inférence causale RB n'est jamais mise en oeuvre de façon non automatique.

Les membres de la famille TETRAD[6] sont des ensembles de programmes (des *packages*) qui, entre autres tâches, réalisent la tâche d'inférence causale RB. Ils réalisent cette tâche de manière modulaire, au sens où chaque utilisation se compose de l'utilisation d'un programme pour l'identification des indépendances probabilistes relatives et de celle d'un programme qui implémente un algorithme RB d'inférence causale – éventuellement, mais pas nécessairement, PC. Selon les hypothèses qu'il émet à propos les modèles du système qu'il étudie, l'utilisateur choisit tel ou tel programme de l'un et de l'autre type. La modularité des ensembles de programmes TETRAD permet d'utiliser seulement un programme implémentant un algorithme d'inférence RB dans le cas où les indépendances conditionnelles relatives sur l'ensemble de variables considéré sont connues.

Dans la section 5.8 de Spirtes et al. (1993), les auteurs rendent compte d'utilisations effectives de TETRAD ; nous y renvoyons le lecteur intéressé. Pour chaque cas rapporté, il convient de porter attention aux hypothèses émises par les auteurs et aux programmes qui sont utilisés. Signalons enfin que les conclusions que Spirtes, Glymour et Scheines tirent des résultats qu'ils obtiennent ont été sévèrement critiquées[7]. Nous reviendrons plus loin sur certains aspects de cette critique. Avant cela, toutefois, il nous faut nous tourner vers les méthodes d'inférence causale plus traditionnelles, hypothético-déductives, auxquelles l'inférence RB peut être comparée.

3.2 Inférence causale traditionnelle

La première tâche qu'il nous faut mener à bien ici consiste clairement à identifier ce à quoi il convient de comparer l'inférence RB, c'est-à-dire à identifier les méthodes d'inférence plus traditionnelles que les méthodes RB viennent concurrencer. Nous nous y attelons dans la première sous-section. Il apparaît alors que ces méthodes se caractérisent largement par des *principes* d'inférence causale. En conséquence, nous sommes conduits à adopter une démarche inverse de celle que nous avons suivie dans la dernière section.

[6] Il ne nous est pas nécessaire ici de distinguer entre ces différents membres.

[7] Voir en particulier Humphreys et Freedman (1996) et Freedman et Humphreys (1999).

Plus explicitement, nous ne partons pas des procédures pour faire émerger les principes, nous partons des principes et mettons en évidence une procédure qui s'y conforme. Cette procédure fait l'objet de la deuxième sous-section. Des questions relatives à la mise en oeuvre de l'inférence causale sont discutées dans la troisième sous-section.

3.2.1 Identification du comparant

Nous venons d'indiquer que, dans la sous-section qui commence, les méthodes d'inférence causale auxquelles nous comparerons l'inférence RB sont identifiées principalement par les principes qui les sous-tendent. Cette affirmation, toutefois, ne saurait nous tenir lieu de méthode. En d'autres termes, il ne serait pas convenable que la question des principes de l'inférence guide l'identification du comparant quand on a annoncé de la comparaison qu'elle porte en particulier sur ces principes. Ce qui est convenable, en revanche, est de commencer par identifier les domaines d'objets visés par les méthodes d'inférence causale RB et de considérer les méthodes d'inférence traditionnellement à l'oeuvre dans ces domaines. C'est ce que nous faisons ici – et qui précisément nous conduira à faire des principes de l'inférence un élément central de l'identification des méthodes auxquelles les procédures RB seront comparées.

3.2.1.1 Domaines d'objets visés par l'inférence causale RB

En vue d'identifier les domaines d'objets auxquels l'inférence causale RB est adaptée, une première remarque est essentielle : les méthodes RB visent à inférer des causes à partir de probabilités et indépendammment de toute expérimentation ou manipulation. En effet, les données probabilistes dont on infère des connaissances causales grâce aux réseaux bayésiens sont des données probabilistes *d'observation*.

L'impossibilité d'expérimenter – quelles qu'en soient les raisons – est à la fois un obstacle majeur à la recherche des causes et un trait qui semble caractériser les sciences sociales, au sens anglo-saxon de l'expression. Par « sciences sociales », les Anglo-Saxons et nous-mêmes dans la suite de ce texte entendons l'ensemble des études dans lesquelles des méthodes relevant des sciences « dures » sont utilisées pour analyser des phénomènes qui ont à voir avec l'homme. Il ne s'agit donc ni de la sociologie, ni même de l'ensemble que constituent la sociologie et l'économie. Plutôt, *certaines* études en sociologie et en économie relèvent des sciences sociales au sens auquel nous faisons référence. Elles en relèvent ainsi que des études en épidémiologie, psychologie,

linguistique..., et pour autant qu'elles mobilisent de manière essentielle des outils statistiques.

3.2.1.2 Inférence causale en sciences sociales

Dans le domaine des sciences sociales au sens que nous venons de définir, la question de la causalité est souvent abordée comme une question *locale*. Le plus souvent, il s'agit de déterminer s'il existe une relation de cause à effet entre deux variables données, par exemple si l'origine ethnique influence causalement la réussite scolaire[8]. A l'inverse, l'inférence causale RB telle que nous l'avons présentée constitue une réponse à la question *globale* des relations de cause à effet au sein d'un ensemble de variables.

Cette différence importante n'implique pas que la comparaison de l'inférence causale RB avec l'inférence causale telle qu'elle se pratique traditionnellement en sciences sociales n'a pas de sens. Au contraire, elle nous permet d'identifier plus précisément ces méthodes d'inférence causale utilisées dans les sciences sociales avec lesquelles il convient de comparer les méthodes RB. En effet, si les études de sciences sociales abordent majoritairement la question de la causalité comme une question locale, certaines d'entre elles l'abordent comme une question globale. Ces études utilisent les méthodes de ce qu'il est convenu d'appeler « modélisation d'équations structurelles » (*structural equation modeling*) ou, plus simplement, « modélisation causale » (*causal modeling*). C'est avec l'inférence causale telle qu'elle se pratique dans le cadre de la modélisation causale que l'inférence RB doit être comparée.

Dans le cadre de la modélisation causale, la question de la causalité au sein d'un ensemble de variables causalement suffisant relève de ce que Kline appelle « analyse de chemins » (*path analysis*). Il convient ici d'être clair : « analyse de chemins » n'a pas alors le sens qui lui est historiquement donné par Wright[9], mais un sens moins précis. En ce sens élargi, l'analyse de chemins se caractérise d'abord par ceci qu'elle s'applique à des ensembles de variables causalement suffisants.

A ce point, il semble que nous avons identifié ce à quoi l'inférence causale RB doit être comparée. Il s'agit de l'inférence causale telle qu'elle est menée dans le cadre de l'analyse de chemins (« inférence causale AC » dans la suite). Le problème est que cette description ne suffit pas à caractériser une procédure, ou même un ensemble de procédures. Plus précisément, on rencontre (au moins) trois obstacles au moment de décrire les procédures

[8]Felouzis (2003).

[9]Wright (1921) est l'article séminal. Voir aussi Wright (1934).

d'inférence causale AC. Le prochain paragraphe est consacré à exposer ces obstacles et à expliquer comment nous les contournons.

3.2.1.3 Obstacles à la description des procédures d'inférence causale AC

En premier lieu, si les sciences sociales telles que nous les avons caractérisées ont l'unité méthodologique impliquée par la commune impossibilité d'expérimenter, il n'est pas aisé de se repérer dans la méthodologie des sciences sociales, ni même dans celle de l'analyse de chemins. En effet, la diversité des objets et, surtout, des domaines institués dont relèvent les sciences sociales ont pour corrélats un flottement dans la terminologie méthodologique même. Ainsi, le même procédé peut être désigné différemment par un épidémiologue et par un sociologue. Il nous semble même que les mêmes termes désignent parfois des procédés différents. Dès lors, nous utilisons le moins de termes techniques possibles et veillons à expliquer en quel sens nous les utilisons. De manière générale, nous reprenons la terminologie utilisée dans Kline (1998).

En deuxième lieu, et de façon apparemment plus problématique, nous rencontrons l'obstacle constitué par ceci que l'objet premier de l'analyse de chemin n'est pas l'inférence causale. Ainsi, l'analyse de chemins ne vise pas d'abord à déterminer si une variable en cause une autre, mais à quantifier l'effet d'une variable sur une autre variable qu'elle cause.[10] Dans les termes que Wright utilise pour décrire la méthode originale :

> . . . une méthode pour mesurer l'influence directe le long chaque chemin pris isolément [...] et donc trouver le degré auquel la variation d'un effet donné est déterminée par chaque cause particulière.[11]

Cependant, cela n'implique pas qu'il soit impossible de définir une procédure d'inférence causale mobilisant les outils de l'analyse de chemins. Plus, nous soutenons que les outils de l'analyse de chemins sont effectivement utilisés pour inférer des causes. Une fois que nous aurons expliqué comment cela se fait, il apparaîtra que c'est même là une utilisation assez naturelle de ces outils. De façon plus générale, il apparaîtra que ce deuxième obstacle et le troisième se contournent d'un même mouvement.

En troisième lieu, le projet de décrire une procédure d'inférence causale AC se heurte à la diversité des pratiques réelles. Cette diversité conduit à envisager qu'il n'existe rien de tel qu'*une* procédure d'inférence causale AC. Toutefois, conclure de la diversité des pratiques à l'absence d'une unité

[10] Ce point est souligné dans Freedman (1987) p. 112.
[11] Wright (1921) p. 557.

méthodologique est aller trop vite en besogne. Positivement, nous soutenons qu'une unité est sous-jacente à la diversité des pratiques. Cette unité est plus précisément celle qu'impose la primauté de la spécification de modèles dans le domaine de la modélisation causale. En d'autres termes, l'unité méthodologique des pratiques dans le domaine de la modélisation causale correspond exactement à ceci que toute analyse relevant de ce domaine commence par la construction d'un graphe orienté dont on considère qu'il représente adéquatement la structure causale du phénomène considéré. Pour ce qui est plus précisément de l'*inférence* causale, cette unité méthodologique est exactement celle que décrit le terme « théorique » tel qu'il est défini plus haut[12] et correspond au caractère hypothético-déductif de l'inférence.

Dans cette optique, la diversité des pratiques réelles se comprend en termes d'écarts à la norme méthodologique de l'hypothético-déduction. Autrement dit, les pratiques réelles diffèrent parce qu'elles s'écartent du canon méthodologique hypothético-déductif, et dans la mesure où elles s'en écartent. Dans ces conditions, il n'est plus seulement impossible de comparer l'inférence causale RB à la *pratique* de l'inférence causale AC ; il n'y a pas grand sens à le faire. Ce à quoi l'inférence causale RB doit être mesurée est ce qu'est l'inférence causale AC si elle ne s'écarte pas des préceptes méthodologiques hypothético-déductifs. Plus précisément, les procédures d'inférence causale RB doivent être comparées à une procédure d'inférence causale hypothético-déductive recourant aux outils de l'analyse de chemins. La prochaine sous-section est consacrée à mettre au jour une telle procédure.

3.2.2 Procédure d'inférence causale traditionnelle

L'objet de la section qui commence est de définir une procédure d'inférence causale hypothético-déductive qui utilise les outils de l'analyse de chemins. En vue de cela, nous commençons par rappeler ce qu'il faut entendre précisément par « hypothético-déduction ».

3.2.2.1 Hypothético-déduction

La définition classique de l'hypothético-déduction est proposée dans Popper (1934) : il s'agit de

> la méthode qui consiste à mettre les théories à l'épreuve dans un esprit critique et à les sélectionner conformément aux résultats des tests, suit

[12]Plus exactement, c'est une caractérisation de « a-théorique » qui a été donnée plus haut. Toutefois, il est clair que cette définition implique une définition de « théorique ».

> toujours la même démarche : en partant d'une nouvelle idée, avancée à titre d'essai et nullement justifiée à ce stade – qui peut être une prévision, une hypothèse, un système théorique ou tout ce que vous voulez –, l'on tire par une déduction logique des conclusions. L'on compare alors ces conclusions les unes aux autres et à d'autres énoncés relatifs à la question de manière à trouver les relations logiques (telles l'équivalence, la déductibilité, la compatibilité ou l'incompatibilité) qui les unissent.[13]

Ainsi, les inférences hypothético-déductives se déroulent en quatre temps :

i) une hypothèse est formulée ;

ii) des conséquences en sont tirées ;

iii) l'hypothèse est mise à l'épreuve ;

iv) l'hypothèse est rejetée ou acceptée.

Pour ce qui est de la mise à l'épreuve de l'hypothèse, Popper distingue en particulier deux choses : d'une part les tests portant sur les conséquences tirées de l'hypothèse[14], d'autre part « la comparaison de la théorie [c'est-à-dire de l'hypothèse] à d'autres théories, dans le but principal de déterminer si elle constituerait un progrès scientifique au cas où elle survivrait à nos divers tests »[15]. Il nous semble clair que la logique de l'hypothético-déduction n'impose pas un ordre dans lequel les différentes mises à l'épreuve devraient avoir lieu.

Maintenant que nous savons précisément ce qu'est l'hypothético-déduction, nous pouvons définir une procédure d'inférence causale hypothético-déductive mobilisant les outils de l'analyse de chemins. Nous le faisons en deux temps. D'abord, nous énumérons les grandes étapes de cette procédure. Ensuite, nos explicitons ce qu'il s'agit de faire à chaque moment.

3.2.2.2 Étapes de l'inférence aux causes traditionnelle

Etant donnés un ensemble de variables $\mathbf{V}$ et des données probabilistes relatives à $\mathbf{V}$, la procédure d'inférence causale AC que nous définissons consiste à :

Étape A : Spécifier un modèle causal sur-identifié (*over-identified*), M.

Étape B : Estimer la valeur des paramètres associés aux différentes relations de cause à effet représentées par M.

Étape C : Tester M et décider s'il doit être rejeté.

[13] Popper (1934) pp. 28–29.

[14] Ces tests peuvent eux-mêmes être de plusieurs types.

[15] Popper (1934) p. 29.

Étape D : Réitérer les étapes A à C pour des modèles différents de M.
Étape E : Identifier celui des modèles non rejetés à l'issue de C qui a la meilleure adéquation (*fit*) aux données, et noter M^* ce modèle.
Étape F : Identifier, parmi des modèles équivalents à M^*, celui dont il est le plus plausible qu'il représente adéquatement la structure causale sur $\mathbf{V}$, et noter $MI_{\mathbf{V}}$ ce modèle.
$MI_{\mathbf{V}}$ est le résultat de la procédure d'inférence causale.

En vue de montrer que cette procédure est bien hypothético-déductive au sens popperien du qualificatif, il nous faut rendre plus sensible et explicite ce en quoi consiste chaque étape de la procédure que nous venons de décrire.

3.2.2.3 Explicitation

Dans le paragraphe qui commence, nous reprenons chacune à son tour les différentes étapes de la procédure d'inférence causale qui vient d'être définie et, pour chacune, nous explicitons ce qui exige de l'être.

Spécifier un modèle causal, c'est définir un graphe orienté qui pourrait représenter adéquatement les relations de cause à effet directes sur l'ensemble de variables $\mathbf{V}$ qu'on considère. Un modèle est sur-identifié si et seulement si il a des degrés de liberté, c'est-à-dire si et seulement si le nombre de ses paramètres est inférieur au nombre d'observations autorisé par $\mathbf{V}$. Une observation autorisée par $\mathbf{V}$ est soit la variance d'une variable de $\mathbf{V}$, soit la covariance entre deux variables de $\mathbf{V}$. Le nombre d'observations autorisées par $\mathbf{V}$ est $\|\mathbf{V}\|(\|\mathbf{V}\|+1)/2$. Pour ce qui est des notions d'identification autres que la sur-identification et des critères d'identification, nous renvoyons le lecteur à Kline (1998).[16]

Estimer les paramètres causaux d'un modèle, c'est déterminer quel est l'effet quantitatif de la variation de la valeur d'une cause supposée sur la valeur de l'un de ses effets supposés. Dans le cas linéaire, le plus simple, c'est estimer quelle différence cela fait sur la valeur de l'effet que la valeur de la cause augmente d'une unité. Pour un modèle sur-identifié, il est théoriquement possible de dériver une estimation unique de chacun des paramètres causaux. Deux options sont possibles pour l'estimation des paramètres causaux d'un modèle :

– l'estimation par régression multiple. Le principe est alors le suivant : pour chaque variable, on considère les relations pour lesquelles elle est effet et on attribue aux paramètres correspondant à ces relations la valeur qui minimise la distance entre les valeurs de V qu'on observe et les valeurs de V que le modèle prédit ;

[16] Kline (1998) pp. 105–110.

– l'estimation par maximum de vraisemblance. Il s'agit alors de maximiser la vraisemblance de l'hypothèse selon laquelle les observations données sont tirées de la population considérée.

Le choix de l'une ou de l'autre de ces options dépend en particulier des hypothèses qu'on émet à propos du modèle.

Tester un modèle, c'est déterminer s'il reste plausible après que les paramètres ont été estimés. Plus précisément, c'est répondre à la question, *fermée*, de savoir si l'hypothèse selon laquelle le modèle considéré représente adéquatement les relations de cause à effet sur $\mathbf{V}$ ne se révèle pas incohérente à la lumière de l'estimation des paramètres causaux. Nous identifions trois types de tests qui peuvent (et donc devraient) être menés à l'étape 3. de la procédure décrite plus haut. Ils consistent respectivement à :

a. s'assurer que les signes et valeurs absolues des estimations obtenues pour les paramètres sont plausibles. Cette vérification doit être à la fois locale et globale. Localement, il s'agit de vérifier que le signe et la valeur absolue de l'estimation de chaque paramètre fait sens. En particulier, chacun de ces paramètres doit être significativement différent de zéro. En effet, si ce n'était pas le cas, alors le modèle – qui précisément postule l'existence de relations de cause à effet auxquelles les paramètres sont associés – doit être rejeté. De façon générale, les conséquences des examens locaux aussi bien que globaux menés à ce point portent toujours sur le modèle lui-même, qui est rejeté ou non ;

b. calculer les résidus de corrélation – c'est-à-dire les différences entre les corrélations impliquées par le modèle et les corrélations observées – et vérifier qu'aucun n'a une valeur absolue supérieure à 0,1[17]. Ce test repose sur ceci que si un modèle est causal, alors la corrélation entre deux variables doit être égale à la somme des paramètres causaux et des corrélations non causales qui figurent le long des différents chemins entre les deux variables. C'est cette somme qu'on appelle « corrélation impliquée par le modèle », et qu'on compare à la corrélation effectivement observée. Dans le cas où le modèle est acyclique, l'égalité entre les corrélations impliquées par le modèle et les corrélations observées est connue sous le nom de « règle du tracé » (*tracing rule*)[18] ;

c. calculer les restrictions de sur-identification (*over-identification restrictions*) – c'est-à-dire la différence entre deux estimations différentes des mêmes paramètres structurels – et vérifier que l'hypothèse selon laquelle elles sont nulles ne peut pas être rejetée. L'idée est ici la suivante : pour des modèles sur-identifiés, il est parfois possible d'estimer

[17] Il s'agit d'une valeur conventionnelle mais généralement acceptée.

[18] Concernant cette règle, voir en particulier Freedman (1987) pp. 112–114.

un même paramètre de plusieurs façons différentes. Si le modèle est correct, ces différentes méthodes doivent donner des résultats identiques. Réciproquement, si ces méthodes donnent des résultats différents, alors le modèle peut être rejeté.

Mesurer l'adéquation d'un modèle aux données, c'est évaluer le degré auquel le modèle est capable de reproduire les données. Il ne s'agit donc plus d'examiner si le modèle en tant qu'il est estimé est plausible, mais de confronter le modèle aux données. En outre, la question posée n'est plus une question fermée, qui requiert une réponse catégorique, mais une question *ouverte*, qui appelle une réponse graduée, quantifiée. C'est précisément ce caractère gradué qui rend possible la comparaison des modèles entre eux sur le critère de leur adéquation aux données. Il existe différentes mesures de l'adéquation d'un modèle aux données, qui portent sur des aspects différents du rapport entre les données et le modèle considéré. Plusieurs de ces grandeurs peuvent être calculées à l'étape E de la procédure d'inférence causale AC. Le « chi-deux » d'un modèle est sans doute la plus fondamentale d'entre elles, au sens où les mesures plus complexes mobilisent presque toujours cette grandeur. Elle représente ce que le modèle explique des corrélations entre les variables considérées. On notera que mesurer l'adéquation d'un modèle n'a de sens que si ce modèle est sur-identifié ; dans le cas contraire, le modèle estimé ne peut qu'être complètement adéquat aux données.

Engendrer un modèle équivalent à un modèle donné, c'est construire un modèle différent du modèle initial mais qui prédit les mêmes corrélations que lui. Des modèles équivalents ont la même adéquation aux données, pour toutes les mesures d'adéquation. Il existe des algorithmes d'engendrement de modèles équivalents à un modèle donné.[19]

Cette explication des termes mobilisés à l'occasion de la description de la procédure d'inférence causale AC achève notre description de cette procédure. Surtout, elle nous permet de montrer en quoi cette procédure est conforme au principe de l'hypothético-déduction tel qu'il est formulé par Popper.

3.2.2.4 Caractère hypothético-déductif de la procédure définie

Il est clair que l'étape A est le moment où est formulée une hypothèse, que l'étape B est le moment où des conséquences sont tirées de cette hypothèse et l'étape C le moment où ces conséquences sont utilisées pour tester l'hypothèse. Plus précisément, l'étape C consiste à déterminer si les conséquences du modèle sont compatibles avec les données. Il nous semble que ce qui se

[19]Pour une présentation, voir Kline (1998) pp. 153–156.

passe après qu'on a testé le modèle estimé est moins immédiatement clair. Nous nous y arrêtons donc plus longuement.

Une fois que le modèle a été testé (à l'étape C), il est rejeté si et seulement si ses conséquences ne sont pas compatibles avec les données. Puis, seulement ensuite et d'abord à l'étape E, a lieu la comparaison entre hypothèses que prescrit Popper. Plus précisément, la comparaison est alors entre les modèles qui n'ont pas été rejetés à l'issue de l'étape C de test. Il apparaît alors clairement que mener la comparaison entre modèles *après* que les modèles ont été testés séparément permet de comparer entre eux seulement des candidats déjà sérieux au titre de modèle adéquat.

Ces candidats sérieux sont comparés sur le critère de leur adéquation aux données. Plus exactement, on retient à l'issue de l'étape E celui des modèles comparés qui a la meilleure adéquation aux données. Une fois ce modèle identifié, on compare les modèles qui lui sont équivalents et on retient le plus plausible d'entre eux. Prises en ensemble, les étapes E et F constituent une inférence à la meilleure explication au sens où Harman définit cette notion : « on infère, de la prémisse qu'une hypothèse donnée fournirait une "meilleure" explication que n'importe quelle hypothèse, la conclusion selon laquelle l'hypothèse donnée est vraie »[20]. D'une part, en effet, l'adéquation d'un modèle aux données est généralement considérée comme sa capacité à expliquer les données. L'étape E vise donc à choisir celle qui *explique le plus* parmi les hypothèses formulées (à l'étape A) et qui n'ont pas été rejetées (à l'issue de C). D'autre part, l'étape F vise à identifier celle qui est la plus plausible, et en ce sens *la meilleure*, parmi des hypothèses également explicatives. En effet, étant équivalentes, elles ont le même score pour toutes les mesures d'adéquation aux données.

3.2.3 Mise en oeuvre de l'inférence causale AC

La plupart des tâches mentionnées dans la procédure d'inférence causale AC ne peuvent être menées à bien qu'en recourant à des programmes informatiques. C'est le cas en particulier pour l'estimation des paramètres d'un modèle à partir de données probabilistes, pour certaines des vérifications imposées par les tests de type a., pour les calculs requis par les tests de type b. et c., pour l'évaluation de l'adéquation d'un modèle aux données sous différentes mesures. Dans ces conditions, la procédure d'inférence causale dans son ensemble peut être menée à bien étape après étape en recourant à un programme convenable à chaque étape qui le requiert.

Sous la modalité que nous venons d'envisager, la mise en oeuvre de

[20]Harman (1965) p. 89.

l'inférence causale AC diffère de la mise en oeuvre de l'inférence causale RB par les programmes TETRAD par un aspect fondamental. Cet aspect est le suivant : les programmes TETRAD sont automatiques au sens où ils construisent un modèle causal à partir des seules données probabilistes. Le caractère automatique de la mise en oeuvre apparaît alors comme le pendant du caractère a-théorique de la stratégie inférentielle. A l'inverse, nous venons d'envisager de mettre en oeuvre l'inférence causale AC sur un mode non automatique : le chercheur intervient à chacun des moments de la mise en oeuvre. Ce mode de mise en oeuvre est celui auquel conduit naturellement le caractère théorique de la stratégie d'inférence AC. Dans ces conditions, il est celui auquel il est le plus fréquemment recouru.

En vue d'étendre le domaine de la comparaison possible entre l'inférence causale RB et l'inférence causale AC, il convient que nous donnions l'intuition de ce qu'automatiser la recherche des causes peut vouloir dire dans le contexte de l'analyse de chemins. Ainsi qu'il est peut-être déjà clair, l'automatisation de l'inférence causale AC est en fait l'automatisation de l'étape A de la procédure que nous avons décrite. Pour être plus précis, la stratégie la plus courante est la suivante : étant donné le modèle vide (dans lequel ne figure aucune flèche) sur l'ensemble de variables considéré, on ajoute la flèche dont la présence augmente le plus l'adéquation du modèle aux données, puis la flèche dont la présence – en plus de la première – augmente le plus l'adéquation du modèle aux données, et ainsi de suite jusqu'à ce qu'un critère d'arrêt de la procédure soit atteint. Il existe des ensembles (*packages*) de programmes qui intègrent des programmes réalisant ce type de tâches. A ces ensembles appartiennent également des programmes menant à bien les tâches classiques dans le domaine de la modélisation causale, et donc en particulier celles dont nous avons vu qu'elles sont impliquées par la procédure que nous avons définie. LISREL constitue la plus ancienne et la plus connue des familles de tels programmes. De même que les ensembles de programmes TETRAD et au même titre qu'eux, les ensembles de programmes LISREL sont modulaires : pour chaque tâche, l'utilisateur peut choisir comment elle devra être réalisée.

Dans la section qui s'achève, nous avons identifé comme relevant de l'analyse de chemins les méthodes d'inférence causale avec lesquelles les méthodes RB peuvent être comparées. Nous avons expliqué que l'inférence causale AC est hypothético-déductive, puis défini une procédure d'inférence causale hypothético-déductive mobilisant les outils de l'analyse de chemins. Pour finir, nous avons été amenés à distinguer entre deux modalités de mise en oeuvre de l'inférence causale AC : automatique ou non automatique. Cela ayant été fait, nous pouvons en venir à la comparaison qui est l'objet principal du présent chapitre.

3.3 Comparaison : Inférence causale RB et et inférence causale traditionnelle

En vue d'organiser notre comparaison de l'inférence causale RB et de l'inférence causale AC, il convient de rappeler que le présent chapitre concerne la méthodologie de l'inférence aux causes. Pour le philosophe, cette question est d'abord celle de la nature de l'inférence causale. Sur ce point, il est déjà apparu que l'inférence RB est déductive là où l'inférence AC est hypothético-déductive. Il nous reste à préciser encore cette distinction et à en expliciter les conséquences. Nous le faisons dans la première sous-section. A la suite de cette première analyse, nous chercherons à intégrer à la comparaison des limites à l'inférence aux causes qui se présentent comme spécifiques de l'une et de l'autre des deux méthodes d'inférence que nous considérons. Ces limites sont discutées dans la deuxième sous-section pour celles qui se présentent comme spécifiques de l'inférence causale AC, et dans la troisième pour celles qui se présentent comme spécifiques de l'inférence causale RB. Dans une quatrième sous-section, nous en revenons à la comparaison et résumons nos résultats.

3.3.1 Comparaison des principes

Si nous avons montré que l'inférence RB et l'inférence AC sont respectivement déductive et hypothético-déductive, les deux qualificatifs n'ont pas été définis exactement au même plan. D'un côté, en effet, le caractère déductif de l'inférence causale RB a été défini relativement au rapport logique entre les prémisses et la conclusion de l'inférence causale. Sur ce plan, le qualificatif « déductif » désigne le caractère nécessaire de ce rapport, le fait que la conclusion de l'inférence est nécessairement vraie si les prémisses le sont. L'inférence RB est déductive en ce sens. De l'autre côté, l'inférence AC est hypothético-déductive en tant qu'elle se conforme à la stratégie d'inférence théorique recommandée par Popper. La question qui se pose à ce point est donc celle de savoir si l'inférence AC est déductive au sens où l'est l'inférence causale RB, c'est-à-dire si la conclusion d'une inférence AC est une conséquence logique de ses prémisses. Nous traitons cette question maintenant.

3.3.1.1 Déduction et hypothético-déduction

Dans ce paragraphe, nous montrons que l'inférence causale AC n'est pas déductive au sens logique où l'est l'inférence causale RB. Plus précisément, nous montrons que le modèle qui constitue la conclusion d'une inférence AC n'est pas nécessairement adéquat si sont vraies les données traitées qui

constituent les prémisses de l'inférence. Par « données traitées », nous entendons l'ensemble des indépendances probabilistes relatives pour les inférences RB et, pour les inférences AC, l'ensemble des corrélations partielles. Ainsi, ce que nous cherchons à montrer est que les corrélations entre les variables d'un ensemble $\mathbf{V}$ n'impliquent pas que les flèches qui figurent dans la sortie $MI_{\mathbf{V}}$ de la procédure AC représentent adéquatement des relations de cause à effet. Nous présentons deux arguments en faveur de cette thèse, le premier générique et le second plus spécifique.

Argument générique. En un sens, que la conclusion d'une inférence causale AC n'est pas une conséquence nécessaire de ses prémisses découle directement de son caractère hypothético-déductif. En effet, la principale caractéristique logique de l'hypothético-déduction est que la conclusion de l'inférence a seulement « provisoirement réussi son test : nous n'avons pas trouvé de raisons de l'écarter »[21].

Plus précisément, la conclusion d'une inférence hypothético-déductive dépend des hypothèses qui ont été envisagées. Dans le cas des inférences AC, elle dépend des modèles causaux que le chercheur qui pratique l'inférence a été capable de spécifier à l'étape A. Elle ne dépend donc pas des seules données traitées et, *a fortiori*, n'en est pas une conséquence nécessaire. Cette caractéristique des inférences AC est reconnue comme une de leurs faiblesses logiques[22]. Elle signifie que les inférences causales AC ne sont pas déductives au sens où le sont les inférences causale RB.

Nous venons de rappeler que, de manière générale, les inférences hypothético-déductives ne sont pas telles que leur conclusion est une conséquence nécessaire de leurs prémisses. Ce que nous voulons montrer maintenant est que, en-deçà de ce point général, il existe des raisons spécifiques pour lesquelles la conclusion d'une inférence causale AC n'est pas une conséquence nécessaire de ses prémisses. En d'autres termes, nous en venons à des caractéristiques propres à l'inférence causale AC qui font qu'elle n'est pas déductive au sens où l'est l'inférence RB.

Argument spécifique. A cet endroit, nous concentrons notre attention sur les étapes E et F. Nous avons vu qu'elles constituent ensemble une inférence à la meilleure explication. Or, rien ne garantit la vérité de l'explication la meilleure. Cette limite méthodologique de l'inférence à la meilleure

[21] Popper (1934) p. 29.

[22] Voir en particulier Freedman (1987) pp. 120–121 et Freedman (1991) pp. 303–304 et 309.

explication est plus précisément corrélative de la difficulté qu'il y a à définir ce que c'est, pour une explication, qu'être meilleure qu'une explication concurrente.[23]

Relativement à l'étape E, cette difficulté prend une forme particulièrement aiguë : « il existe des douzaines de mesures d'adéquation des modèles aux données [...] et de nouvelles mesures sont développées en permanence »[24] qui, parce qu'elles ne visent pas exactement le même aspect du rapport entre modèle et données, n'ordonnent pas les différents modèles de la même façon. Pour ce qui est de l'étape F, il s'agit plus simplement de ceci qu'il est difficile de définir précisément l'idée de plausibilité d'une hypothèse.

De manière générale, il suit de ce que les étapes E et F constituent une inférence à la meilleure explication, que le modèle qu'elles sélectionnent n'est pas nécessairement adéquat à la structure causale réelle. Plus précisément, dans chaque cas, le critère de sélection d'un modèle parmi des modèles concurrents n'est pas tel que le modèle causal adéquat est nécessairement sélectionné s'il figure parmi ceux entre lesquels il s'agit de discriminer.

Il apparaît donc, au total, que les étapes E et F d'une inférence AC ne sont pas telles que la conclusion est une conséquence logique des prémisses. Notons bien que l'analyse qui aboutit à cette conclusion porte sur les *seules* étapes E et F. En d'autres termes, les raisons de non-déductivité de l'inférence AC qui nous venons de mettre au jour sont bien attachées à la spécification AC du précepte popperien de comparaison des modèles. L'analyse compte alors sans ceci que l'ensemble des modèles entre lesquels il s'agit de discriminer à l'étape E dépend de l'ensemble des modèles spécifiés en A.

En définitive, il est apparu deux raisons sensiblement différentes pour lesquelles la conclusion d'une inférence causale AC n'est pas une conséquence nécessaire des données probabilistes d'observation traitées qu'elle prend pour prémisses. La première raison est générale, tenant à la logique même de l'hypothético-déduction. La seconde raison est spécifique de l'inférence causale AC, et tient précisément à la façon dont les outils de l'analyse de chemins permettent de donner corps au précepte popperien de comparaison des hypothèses. La suite de la sous-section est consacrée à examiner les conséquences de ce trait propre aux inférences RB qu'est leur déductivité.

3.3.1.2 Conséquences de la déductivité

Nous venons de montrer que les inférences causales RB se distinguent des inférences causales plus traditionnelles qu'elles peuvent venir concurren-

[23] Harman (1965) p. 89.
[24] Kline (1998) p. 133.

cer par leur caractère déductif. En vue d'en examiner les conséquences de ce résultat, il est très utile de revenir sur l'argument générique en faveur de la thèse selon laquelle les inférences causales AC ne sont pas déductives. En effet, l'argument consiste en particulier à faire valoir que le caractère théorique de la stratégie inférentielle AC implique que les inférences AC ne sont pas déductives. Plus explicitement, le fait que l'inférence procède de la formulation d'hypothèses implique que la conclusion de l'inférence n'est pas une conséquence nécessaire de ses prémisses. Par contraposition, il apparaît alors que le caractère déductif de l'inférence causale RB implique que la stratégie inférentielle est a-théorique. En d'autres termes, une condition nécessaire pour que les réseaux bayésiens fondent une stratégie d'inférence a-théorique est que les inférences causales RB soient déductives. La déductivité des inférences RB est logiquement inséparable du caractère a-théorique de la stratégie inférentielle. Dans ces conditions, on voit mal comment on pourrait isoler l'effet de la nature spécifique des inférences RB – c'est-à-dire de leur déductivité – de celui de leur a-théoricité.

Toutefois, il existe un moyen de séparer l'effet de la déductivité et l'effet de l'a-théoricité sur l'inférence aux causes. Pour le comprendre, il convient de se pencher sur les questions de mise en oeuvre de l'inférence. Nous avons vu dans les sous-sections 3.1.3 et 3.2.3 que l'a-théoricité a pour pendant l'automatisation de la recherche des causes. A l'inverse, l'inférence causale théorique est mise en oeuvre le plus naturellement d'une manière non automatique. Mais nous avons vu aussi dans la sous-section 3.2.3 qu'il existe des méthodes automatiques pour l'inférence causale AC : les ensembles de programmes de la famille LISREL. Ces programmes, étant automatiques, ne suppose pas qu'une hypothèse théorique soit formulée par le chercheur qui mène l'inférence. Du coup, en s'intéressant aux inférences causales menées grâce à LISREL, on isole le caractère non déductif des inférences AC de leur caractère théorique. La comparaison de ces inférences à celles qu'on mène grâce à TETRAD permet donc d'évaluer l'effet net de la déductivité des inférences RB. Nous menons cette comparaison dans la fin du présent paragraphe.

La comparaison des inférences causales menées grâce à TETRAD aux inférences causales menées grâce à LISREL est largement en faveur de TETRAD.[25] De cela, il existe plusieurs raisons non indépendantes. Parmi celles-ci, nous en présentons deux qui résultent clairement de ce que les inférences RB sont déductives la où les inférences AC ne le sont pas.

En premier lieu, TETRAD n'a pas pour résultat un modèle (comme c'est le cas de LISREL), mais un patron représentant une classe de modèles

[25]Voir par exemple Spirtes et al. (1990).

équivalents. Il s'agit d'un avantage de TETRAD, présenté comme tel dans Spirtes et al. (1993)[26]. En effet, dans le contexte de recherche *automatique* des causes, il n'y a pas de raison de préférer à un autre un modèle dont il est indiscernable par les données. Cette caractéristique de TETRAD découle de ce que son résultat n'est ni plus, ni moins que la représentation de toute l'information causale qui peut être *déduite* des données traitées. A l'inverse, LISREL, procédant par spécification puis test de modèles, a pour résultat *un* modèle.

La seconde raison pour laquelle TETRAD est supérieur à LISREL est que la sortie donnée par LISREL dépend du chemin emprunté pour y arriver. Pour le dire simplement, une flèche ajoutée ne peut plus être retirée. Il s'agit clairement d'une limite de LISREL, que ne connaît pas TETRAD. A nouveau, l'avantage de TETRAD tient au caractère déductif de l'inférence RB : quand la conclusion de l'inférence est une conséquence logique de ses prémisses, les voies empruntées pour arriver à la conclusion ne peuvent pas avoir d'importance. De façon générale, il apparaît donc que le caractère déductif des inférences RB implique la supériorité des inférences causales automatiques RB sur les inférences causales automatiques AC.

La comparaison des inférences RB et AC a porté dans cette première sous-section sur les seuls *principes* de l'inférence causale. En particulier, les questions de mise en oeuvre n'ont été abordées que dans la mesure où elles donnent à voir des conséquences de la différence entre une inférence déductive d'un côté et une inférence hypothético-déductive de l'autre. En nous concentrant sur les principes de l'inférence causale, nous n'avons pas pris en compte la question de sa validité. En d'autres termes, nous n'avons pas abordé la question de savoir si l'inférence causale menée selon les procédures RB et AC que nous avons décrites est valide. C'est cette question que nous envisageons maintenant. Plus précisément, la prochaine sous-section est consacrée à discuter des limites à la validité de l'inférence causale qui se présentent comme spécifiques de l'inférence causale AC.

3.3.2 Limites spécifiques de l'inférence causale AC

Dans la dernière sous-section, nous avons montré en particulier que les principes de l'inférence causale AC sont tels que la conclusion d'une inférence causale AC n'est pas nécessairement vraie si les prémisses de l'inférence le sont. Les limites que nous mettons en évidence dans le paragraphe qui commence ont une autre origine. Elles ne découlent pas de la nature même des

[26]Spirtes et al. (1993) pp. 77–78.

inférences AC, mais plutôt de la nature des outils qui sont utilisés pour mener à bien les différentes étapes de la procédure d'inférence causale. Il n'en reste pas moins qu'elles constituent de nouvelles raisons pour lesquelles les inférences causales AC ne sont pas complètement fiables.

Exposées dans l'ordre des étapes de la procédure auxquelles elles sont attachées, ces raisons sont les suivantes :

- à l'étape B, les paramètres associés au modèle considéré ne sont pas *déduits* des données probabilistes d'observation ; ils sont seulement *estimés* à partir d'elles ;
- à l'étape C, les modèles rejetés ne sont pas réfutés au sens strict ; ils sont seulement rejetés parce que la probabilité qu'ils confèrent aux données observées est jugée insuffisamment élevée. Pour être plus précis, les tests mis en oeuvre à l'occasion de l'étape C portent sur des hypothèses *statistiques*. Or, il est bien connu que, pour de telles hypothèses, la réfutation au sens logique laisse la place à un concept méthodologique de réfutation, que la définition de zones critiques guident des décisions de rejet plutôt que des réfutations au sens logique strict. L'idée de réfutation méthodologique des hypothèses statistiques trouve son origine dans le concept popperien de falsifiablité pratique des énoncés probabilistes. En même temps qu'il en introduit le principe, Popper n'a de cesse d'en souligner les limites logiques :

  > Il est clair que cette "falsification pratique" ne peut résulter que de la décision méthodologique de considérer des événements hautement improbables comme exclus – prohibés. Mais de quel droit peut-on les considérer ainsi ? Où devons-nous tracer la ligne de séparation ? Où commence cette "haute improbabilité" ?
  >
  > ... D'un point de vue purement logique, il ne peut y avoir de doute : c'est un fait que les énoncés de probabilité ne peuvent pas être falsifiés.[27]

  Pour ce qui concernent les tests menés à l'étape C de la procédure d'inférence causale AC, il en découle qu'ils peuvent conduire au rejet d'hypothèses correctes ou à l'absence de rejet d'hypothèses incorrectes. Il convient ici d'être clair : ce qui est en jeu n'est pas la simple possibilité que des hypothèses inadéquates passent le test, mais les possibilités – *propres au contexte statistique* – que des hypothèses adéquates échouent à le passer et qu'à l'inverse le passent des hypothèses qui n'ont pas la propriété visée par le test.

De façon plus générale, il apparaît que les limites à la validité de l'inférence causale AC que nous venons de mettre en évidence procèdent toutes deux de

[27] Popper (1934) pp. 192–193.

ceci : les données probabilistes qui constituent les prémisses de l'inférence ne sont pas les corrélations dans la population considérée, mais les corrélations dans un *échantillon* de cette population. Mais alors il apparaît aussi que les difficultés que nous venons de discuter *ne sont pas* spécifiques de l'inférence AC. Revenons, en effet, aux inférences RB. Nous avons montré dans la sous-section 3.3.1 que les principes sur lesquels ces inférences reposent sont tels que la conclusion est une conséquence logique des prémisses probabilistes. Ces prémisses probabilistes consistent en fait en l'ensemble $\mathbf{I_V}$ des indépendances probabilistes relatives entre les variables de l'ensemble $\mathbf{V}$ considéré. Or, si on utilise les réseaux bayésiens pour inférer des causes de données probabilistes d'observations, ces indépendances ne sont pas données. Ce qui est donné alors sont des fréquences relatives et des corrélations au sein d'un échantillon de la population étudiée. Les indépendances probabilistes relatives dans la population sont seulement des hypothèses statistiques qui n'ont pas été rejetées.

Le point qui est le nôtre dans le dernier paragraphe n'est pas nouveau. Il est soulevé en particulier dans Humphreys et Freedman (1996)[28]. Néanmoins, il nous semble qu'il reste encore souvent – trop souvent – négligé. Aussi y insistons-nous : les réseaux bayésiens ne donnent aucun moyen de résoudre les problèmes qui découlent de ce qu'on connaît seulement les fréquences relatives dans un échantillon de la population. Bien qu'en d'autres points peut-être, l'inférence causale RB se heurte à ces problèmes aussi bien que l'inférence causale AC.

Nous voyons deux explications non exclusives l'une de l'autre, ni sans doute même indépendantes, à la négligence dont ce point souffre le plus souvent. D'abord, il nous semble clair que c'est bien pour la nature de l'inférence causale qu'ils autorisent que les réseaux bayésiens renouvellent l'épistémologie de la causalité générique. Ainsi, ce qui distigue l'inférence causale RB dans le champ de l'épistémologie générique est d'abord la déductivité que nous avons mise en évidence et caractérisée dans la sous-section 3.1.2. Ensuite, le débat sur l'inférence aux causes RB s'est focalisé sur les hypothèses sous lesquelles les conclusions de l'inférence causale RB sont des conséquences logiques des indépendances probabilistes relatives au sein de la population considérée. Pour le dire plus clairement, le débat s'est focalisé sur les trois hypothèses corrélatives de la notion de réseau bayésien causal – l'hypothèse d'acyclicité, l'hypothèse de fidélité et la condition de Markov causale – au détriment en particulier du point que nous venons de souligner. Il convient donc que nous nous tournions maintenant vers ces trois hypothèses.

[28]Humphreys et Freedman (1996) p. 117.

3.3.3 Limites spécifiques de l'inférence RB : Acyclicité, condition de Markov causale, fidélité

Ainsi que nous venons de le rappeler, la conclusion d'une inférence causale RB n'est une conséquence logique des indépendances probabilistes relatives au sein de la population considérée qu'aux conditions que le graphe causal sur l'ensemble de variables considéré soit acyclique et que cet ensemble de variables satisfasse la condition de Makov causale et l'hypothèse de fidélité. Or, nous savons depuis le chapitre 1 que l'acyclicité et la condition de Markov causale ne sont pas toujours satisfaites. Par ailleurs, nous avons vu dans le chapitre 2 qu'il en va de même de l'hypothèse de fidélité. Dans un premier temps, nous analysons ce qui en découle pour la validité des inférences causales RB. Dans un second temps, nous examinons la thèse selon laquelle la difficulté rencontrée serait spécifique de l'inférence RB.

3.3.3.1 Conséquences pour l'inférence causale

De ce que la conclusion de l'inférence causale RB n'est une conséquence logique de ses prémisses que sous les trois hypothèses que nous venons de rappeler, il découle en première approche que seuls les ensembles de variables qui satisfont ces hypothèses sont susceptibles d'inférence causale RB. Autrement dit, il semble que l'existence de violations de ces hypothèses limite le domaine dans lequel on peut recourir à l'inférence causale RB. En conséquence, il semble que seuls les ensembles de variables qui satisfont les hypothèses corrélatives de la notion de réseau bayésien causal peuvent jouir du caractère déductif de l'inférence causale RB.

Cette première analyse, toutefois, est trompeuse. Plus précisément, les hypothèses d'acyclicité et de fidélité et la condition de Markov causale ne peuvent pas être considérées comme de simples limites d'un domaine à l'intérieur duquel l'inférence causale pourrait être menée au moyen des réseaux bayésiens et, donc, être déductive. La raison en est simple : *puisque l'inférence causale RB est a-théorique*, l'enquête visant à déterminer avant-coup si les hypothèses nécessaires à la validité de l'inférence sont satisfaites n'a simplement pas d'objet. En effet, avant de mener l'inférence causale RB, nous n'avons pas de graphe causal dont on pourrait se demander s'il est acyclique et s'il est avec les probabilités dans le rapport que décrivent la condition de Markov causale et l'hypothèse de fidélité.

A titre d'explicitation de cette thèse, considérons à nouveau la condition de Markov causale, dont nous avons déjà indiqué qu'elle est celle des trois hypothèses qui a fait couler le plus d'encre. Nous verrons dans le chapitre 4 qu'il existe des résultats de validité de la condition de Markov causale pour

les ensembles de variables qui satisfont certaines propriétés. En particulier, nous verrons que les ensembles de variables déterministes dont les variables exogènes sont conjointement indépendantes satisfont prouvablement la condition de Markov causale. Or, dans le contexte a-théorique qui nous intéresse présentement, ce résultat ne peut pas être utilisé comme un critère *d'identification* d'ensembles de variables qui satisfont la condition de Markov causale. Le critère, en effet, est causal à deux titres : d'une part les variables exogènes d'un ensemble de variables $\mathbf{V}$ sont celles qui n'ont pas de *cause* dans $\mathbf{V}$, d'autre part un $\mathbf{V}$ est déterministe si la valeur des variables non exogènes de $\mathbf{V}$ est déterminée par celle de leurs *causes* directes dans $\mathbf{V}$. Dans ces conditions, décider si un ensemble de variables $\mathbf{V}$ est déterministe et tel que ses variables exogènes sont conjointement indépendantes requiert de connaître le graphe causal sur $\mathbf{V}$. Mais c'est précisément là tout ce que vise l'inférence causale.

De façon plus générale, les trois hypothèses corrélatives de la notion de réseau bayésien causal ont un contenu causal, et il en découle qu'on ne peut pas déterminer si elles sont satisfaites par un ensemble de variables $\mathbf{V}$ sans disposer d'un graphe causal, même hypothétique, sur $\mathbf{V}$. Maintenant, l'absence de connaissance causale n'implique pas seulement qu'il est impossible de déterminer si les trois hypothèses sont satisfaites. Elle implique aussi qu'il est impossible de majorer la déviation entre $GI_{\mathbf{V}}$ et le graphe causal réel sur $\mathbf{V}$ que peuvent occasionner des violations éventuelles des hypothèses. Dans ces conditions, entre la conclusion d'une inférence causale RB et le graphe causal sur l'ensemble de variables considéré, il n'existe aucune forme d'adéquation qui soit garantie. L'hypothèse d'acyclicité, la condition de Markov causale et l'hypothèse de fidélité sont donc plus que les limites du domaine à l'intérieur duquel les inférences causales peuvent être menées grâce aux réseaux bayésiens et jouir de la propriété d'être déductives. Elles sont ce qui fait que l'inférence causale RB n'est jamais fiable – au sens précis où on ne peut jamais être assuré qu'elle donne un résultat correct, même quand c'est le cas.

Avant d'aller plus loin, nous souhaitons souligner ici que par ce que nous venons d'en montrer, les hypothèses corrélatives de la notion de réseau bayésien causal ont un statut différent des hypothèses qui sous-tendent l'utilisation des outils de l'analyse de chemins. Plus clairement, nous avons mentionné plus haut que l'utilisation des différentes méthodes d'estimation des paramètres et des différentes mesures d'adéquation aux données est conditionnée à la satisfaction de certaines hypothèses par le modèle étudié.[29] Ces

[29]Pour une présentation, voir par exemple Kline (1998) ou Kenny (1979). Pour une analyse : Freedman (1987) pp. 101–116, Clogg et Haritou (1997).

hypothèses sont contraignantes et conduisent à des mésusages des outils de l'analyse de chemins. Ces mésusages ont été décrits aussi bien dans les ouvrages méthodologiques[30], que dans la littérature critique[31] ; on a même cherché à les expliquer.[32]

Mais, aussi dommageables ces mésusages soient-ils, leurs conséquences ne sont pas comparables à celles, décrites plus haut, de l'utilisation des procédures d'inférence causale RB dans le cas général. En effet, les hypothèses que ces mésusages consistent à ignorer ne sont pas du même type que l'hypothèse d'acyclité, l'hypothèse de fidélité et la condition de Markov causale. Contrairement à celles-ci, elles ne portent pas sur la causalité, mais seulement sur les distributions de probabilités. Soulignons que même les hypothèses d'indépendance entre termes d'erreur et variables exogènes ne sont pas causales : les variables exogènes ne sont pas les variables exogènes du graphe causal adéquat, mais toujours seulement celles du modèle en train d'être testé. Dès lors qu'elles ne sont pas causales, ces hypothèses peuvent être testées avant d'utiliser ou après avoir utilisé les outils de l'analyse de chemins qui requièrent qu'elles soient satisfaites. La critique adressée à leur propos aux utilisateurs de l'analyse de chemins n'a donc pas le même statut que celle que nous formulions plus haut à l'endroit de l'inférence causale RB. On reproche à ces utilisateurs une négligence pratique, là où nous pointions une difficulté théorique pour l'inférence RB. Nous ne reviendrons pas dans la suite sur les hypothèses probabilistes qui doivent être satisfaites pour qu'il soit possible d'utiliser les différents outils de l'analyse de chemins.

3.3.3.2 Limites spécifiques de l'inférence causale RB ?

Avant d'intégrer l'analyse qui précède à notre comparaison des inférences RB et AC, il convient de nous arrêter à la question de savoir si les difficultés que nous venons de discuter sont bien spécifiques de l'inférence causale RB. Plus clairement, ce paragraphe traite de la question de savoir si l'hypothèse d'acyclicité, l'hypothèse de fidélité et la condition de Markov causale sont bien une caractéristique qui distingue les inférences RB des inférences AC. Sur ce point, le cas de l'hypothèse d'acyclicité est le plus simple. Ainsi, que le graphe causal sur un ensemble de variables donné soit ou non cyclique ne change rien à la possibilité de mener une inférence AC, ni surtout à la pertinence de la conclusion qu'on peut espérer tirer d'une telle inférence. En effet, l'ensemble des tâches que nous avons explicitées dans le paragraphe 3.2.2.3 peuvent être accomplies pour des modèles cycliques – ou, pour reprendre le terme utilisé

[30] Voir par exemple Kline (1998) chap.12.
[31] Voir par exemple Freedman (1991).
[32] Blalock (1991).

dans le domaine de l'analyse de chemins, non-récursifs.

Pour ce qui est maintenant, de la condition de Markov causale et de l'hypothèse de fidélité, elles véhiculent une conception du rapport entre causalité et probabilités dont nous soutenons qu'elle est très similaire à celle qui sous-tend la pratique de l'analyse de chemins. Plus précisément, nous soutenons que la pratique traditionnelle des sciences sociales repose sur une caractérisation de la causalité comme corrélation qui ne disparaît pas par conditionalisation. On considèrera qu'une variable C cause (directement) une variable E si 1) C et E sont dépendantes en probabilités et 2) les autres causes de E ne font pas écran entre C et E.

A l'appui de la thèse selon laquelle on considère comme susceptibles d'être en relation de cause à effet seulement des variables dépendantes en probabilité, nous soutenons d'abord qu'on ne spécifie (à l'étape A de la procédure que nous avons décrite dans la sous-section 3.2.2) un modèle dans lequel figure une flèche entre deux variables données que si ces deux variables sont corrélées. Par ailleurs, nous avons mentionné déjà que si l'estimation d'un modèle donne au paramètre associé à une des flèches du modèle une valeur non significativement différente de zéro, alors ce modèle est rejeté. A ce fait, nous ajoutons ici la précision qu'il est rejeté au profit d'un modèle identique sauf pour ceci que cette flèche n'y figure pas.

A l'appui de la thèse selon laquelle la corrélation entre C et E est interprétée causalement si elle ne disparaît pas par conditionalisation sur les causes de E, nous revenons sur les tests de type b. de l'étape C. Ces tests consistent à vérifier que les corrélations impliquées par un modèle ne diffèrent pas significativement des corrélations observées. Or, les corrélations impliquées par le modèle se calculent par sommation relativement à tous les chemins causaux. Les tests de type b. reposent donc sur l'idée selon laquelle la dépendance entre deux variables qui ne se causent pas l'une l'autre s'épuise dans la dépendance qu'impliquent les différents chemins causaux menant de l'une à l'autre. Réciproquement, une dépendance que n'épuiserait pas la prise en compte de ces chemins demande à être interprétée causalement. Le modèle qui échoue à l'interpréter de cette façon est rejeté à l'issue du test b.

De façon générale, il apparaît que l'inférence causale AC repose sur une conception du rapport entre la causalité et les probabilités qui est très proche de celle que véhiculent les réseaux bayésiens causaux. En ce sens, la condition de Markov causale et l'hypothèse de fidélité ne constituent pas un problème qui serait spécifique de l'inférence causale RB, et qui la distinguerait de l'inférence causale AC.

Toutefois, ce qui précède fait apparaître que ces hypothèses n'ont pas le même statut selon qu'on considère l'inférence causale AC ou l'inférence causale RB. Dans le cas de l'inférence causale AC, nous venons d'identifier les

endroits de la procédure auxquels elle est mobilisée. A l'inverse, le recours à la condition de Markov causale et à l'hypothèse de fidélité n'est pas *localisable* dans l'inférence causale RB. En effet, il en constitue le *principe* même.

Par ailleurs, nous savons maintenant (et depuis la sous-section 3.1.2) que ce principe fonde un mode d'inférence a-théorique. Nous avons montré (dans le dernier paragraphe) que l'a-théoricité implique qu'il est impossible de déterminer avant de mener l'inférence causale RB si la condition de Markov causale et l'hypothèse de fidélité sont satisfaites. Nous en avons déduit que l'inférence causale RB est toujours suspecte de donner des résultats incorrects. De l'autre côté, nous avons vu (à la fin du paragraphe 3.2.2.4) que l'inférence causale AC n'est pas a-théorique, qu'elle repose sur la spécification de modèles causaux. Il en découle qu'on peut tester si un modèle donné satisfait la condition de Markov causale et l'hypothèse de fidélité. Plus précisément, il est possible de procéder à un test statistique de l'hypothèse selon laquelle elles seraient satisfaites dans le cas où le modèle serait correct. Par conséquent, le fait de reposer sur une conception du rapport entre causalité et probabilités du type de celle que véhiculent la condition de Markov causale et l'hypothèse de fidélité n'a pas pour l'inférence AC les conséquences qu'il a pour l'inférence RB. De façon plus générale, il apparaît que si les violations de la condition de Markov causale et de l'hypothèse de fidélité ne sont pas un problème spécifique de l'inférence RB, le statut de ces hypothèses et les conséquences de ces violations le sont.

3.3.4 Inférence causale RB et inférence causale AC : bilan

Il est temps de renouer les fils, et de dresser un bilan ordonné des différentes enquêtes comparatives que nous avons menées depuis le début de la présente section. Pour commencer par le plus évident, il est apparu que ni le problème de la réfutation d'hypothèses statistiques n'est spécifique de l'inférence causale AC, ni le problème constitué par les violations de la condition de Markov causale et de l'hypothèse de fidélité n'est spécifique de l'inférence causale RB. Pour cette dernière raison, on peut considérer que les réseaux bayésiens ont le mérite de rendre explicites des hypothèses sur lesquelles nos inférences causales reposent et qui sont traditionnellement ignorées.

Si les problèmes apparemment spécifiques de l'inférence causale RB ne le sont finalement pas, il ne reste pour caractériser l'inférence causale RB que ceci qu'elle est *déductive*. Par là, il faut entendre que la conclusion d'une inférence causale est nécessairement vraie si le sont les données traitées qui

constituent ses prémisses. Nous avons montré au début de la présente section que cette caractéristique la distingue bien de l'inférence causale AC, qui n'est pas déductive en ce sens.

Il semble clair que la déductivité entendue au sens logique que nous venons de rappeler est une qualité d'une inférence. Nous avons toutefois essayé de mesurer ses effets plus précis dans le cadre de l'inférence aux causes. Pour cela, nous avons été amenés à comparer ce qu'on peut attendre de l'inférence causale automatique RB et ce qu'on peut attendre de l'inférence causale automatique AC. En d'autres termes, nous avons confronté l'inférence causale telle qu'elle est menée par TETRAD à l'inférence causale telle qu'elle est menée par LISREL. La première est apparue supérieure à la seconde : la déductivité profite bien spécifiquement à l'inférence aux causes.

Seulement, la déductivité de l'inférence causale ne vient pas seule. Plus précisément, nous avons vu (au début du paragraphe 3.3.1.2) qu'elle a pour condition nécessaire l'*a-théoricité* – c'est-à-dire le fait de tirer des conclusions générales de prémisses particulières, indépendamment de la formulation d'une hypothèse théorique. C'est parce qu'elle ne procède pas de la formulation d'hypothèses causales que l'inférence RB peut être déductive.

Or l'a-théoricité est précisément la raison pour laquelle la possibilité que l'hypothèse d'acyclicité, la condition de Markov causale et l'hypothèse de fidélité soient violées est rhédibitoire pour l'inférence causale RB. Pour le redire : en l'absence de modèles spécifiés sur la base de considérations théoriques, il n'est jamais possible de s'assurer que les hypothèses sont satisfaites. En conséquence, il est toujours logiquement possible qu'elles soient violées et le résultat d'une inférence RB est toujours suspect de ne pas être correct – même quand il se trouve qu'il l'est. De l'autre côté, précisément parce qu'elles procèdent de la formulation d'hypothèses causales, les inférences AC ne sont pas aussi vulnérables à la possibilité que soient violées la condition de Markov causale et l'hypothèse de fidélité. Si elle n'est pas un problème spécifique de l'inférence RB, cette possibilité a un statut spécifique dans le cadre RB. Encore une fois, elle implique que les conclusions d'une inférence causale RB ne peuvent jamais être dignes de confiance, et ce alors qu'elles peuvent être vraies.

Est-ce à dire que les propriétés remarquables des réseaux bayésiens et les puissants algorithmes d'inférence causale RB ne peuvent aucunement être mis au service de l'inférence causale ? C'est le point que nous discutons dans la dernière section.

3.4 Discussion

Dans la section qui commence, nous discutons le point de savoir si et en quoi les réseaux bayésiens peuvent servir l'inférence causale. Or nous avons montré que le mode d'inférence causale que définissent les algorithmes RB a des limites méthodologiques rhébitoires. Il en découle que la question de l'utilisation des réseaux bayésiens à des fins d'inférence causale est abordée sur le mode prescriptif. Il s'agit essentiellement de déterminer comment les réseaux bayésiens *pourraient* être utilisés pour inférer des causes. Ce que nous envisageons dans cette section sont des propositions méthodologiques relativement à l'inférence causale.

Ce qui précède rend clair qu'utiliser les réseaux bayésiens à des fins d'inférence causale ne peut avoir que le sens suivant : intégrer les algorithmes RB à une procédure hypothético-déductive. Ce que nous allons envisager sont donc des *méthodologies mixtes* d'inférence causale. Mais prendre en compte ce qui précède est aussi prendre en compte le résultat selon lequel l'inférence aux causes traditionnelle repose sur une conception du rapport entre causalité et probabilités qui est similaire à celle que véhiculent la condition de Markov causale et l'hypothèse de fidélité. Plus précisément, c'est prendre en compte la possibilité que ces hypothèses soient violées.

Au total, la présente section vise à proposer une procédure d'inférence causale hypothético-déductive 1) au sein de laquelle les algorithmes RB aient une place et 2) telle que les violations éventuelles de la condition de Markov causale et de l'hypothèse de fidélité puissent être prises en compte – selon des modalités à préciser.

3.4.1 Discussion d'une proposition existante

Il existe une proposition de méthodologie mixte, qui ménage une place aux algorithmes RB au sein de l'inférence causale AC. Selon cette proposition, les algorithmes RB sont mobilisés à l'étape A de la procédure AC. Ils servent donc à formuler des hypothèses causales. La proposition, déjà évoquée dans Glymour et al. (1988)[33], est développée avec le plus de soin dans Williamson (2002). Dans sa version développée, elle est la suivante : utiliser les algorithmes RB pour formuler une hypothèse, déduire de cette hypothèse des prédictions, tester ces prédictions, amender l'hypothèse en fonction du résultat des tests, déduire de la nouvelle hypothèse des prédictions...[34]

Le principal avantage de cette proposition est qu'elle utilise les algorithmes RB pour combler une lacune patente de la méthodologie hypothético-

[33] Glymour et al. (1988) pp. 428–429.
[34] Williamson (2002) pp. 6–7.

déductive. En effet, Popper ne propose pas de méthode pour formuler des hypothèses – considérant d'ailleurs qu'une telle méthode ne saurait exister. En particulier, il ne propose pas de méthode pour formuler des hypothèses causales. En utilisant les algorithmes RB à l'étape A de la procédure d'inférence causale AC, on se donne une telle méthode.

La première limite de cette proposition est précisément qu'elle ne comble pas *complètement* la lacune qu'elle vise à combler. Rappelons, en effet, que le résultat d'un algorithme RB n'est pas un modèle, mais un patron représentant un ensemble de modèles équivalents. Williamson le mentionne, mais considère que la difficulté n'est pas pertinente :

> Ici, les techniques de l'intelligence artificielle sont utilisées pour engendrer un modèle causal (ou un ensemble de modèles causaux, auquel cas plusieurs hypothèses sont évaluées simultanément – j'utiliserai le singulier dans la suite pour des raisons de simplicité) (*for simplicity's sake*).[35]

Nous ne soutenons pas que Williamson ait tort de considérer que plusieurs hypothèses puissent être testées simultanément. Il suffit pour cela de tester une conséquence qui est commune à ces hypothèses. Toutefois, le fait que les algorithmes RB ont pour résultats des patrons, et non des graphes orientés acycliques, interdit de leur attribuer dans l'inférence causale le rôle envisagé par Williamson. Si, d'un côté, la conclusion de l'inférence causale doit bien être *un* modèle causal et si, de l'autre côté, l'inférence causale doit bien être hypothético-déductive, alors il faut à un moment ou l'autre formuler une hypothèse causale qui est *un* modèle causal. En conséquence, les algorithmes RB d'inférence causale ne peuvent pas être le seul outil de spécification d'hypothèses causales – contrairement à ce que Williamson laisse entendre. La spécification d'hypothèses causales repose toujours *in fine* sur des considérations théoriques et, du coup, on ne discerne plus très bien le gain méthodologique associé à cette proposition.

En outre, la proposition de Williamson souffre à nos yeux de ce qu'elle ne prend pas en compte le problème que constitue pour l'inférence causale AC la possibilité que soient violées la condition de Markov causale et l'hypothèse de fidélité. D'une part, tels qu'ils sont intégrés à l'inférence causale AC, les algorithmes RB sont utilisés de la manière aveugle dont nous avons montré dans la sous-section 3.3.3 qu'elle est méthodologiquement problématique. D'autre part, Williamson ne prend pas en compte ceci que la condition de Markov causale et l'hypothèse de fidélité sont sous-jacentes à la pratique et supposées par certains des outils de l'inférence causale AC. Signalons que ce point ne peut pas être opposé à Williamson. En effet, sa proposition est formulée à

[35] Williamson (2002) p. 7.

un niveau d'abstraction plus élevé que celui auquel nous nous plaçons dans le présent chapitre, et auquel la nature des outils de l'analyse de chemins importe peu. Il n'en reste pas moins que ce point constitue une raison pour nous ici de rejeter l'idée selon laquelle les algorithmes RB pourraient simplement servir à formuler des hypothèses causales.

3.4.2 Formulation d'une nouvelle proposition

Si les algorithmes RB ne peuvent pas être utilisés au moment de formuler les hypothèses causales, il reste à envisager de les utiliser pour tester des hypothèses. Sous sa spécification la plus naturelle, la proposition est d'utiliser les algorithmes RB pour fonder un nouveau test qui viendrait prendre place à l'étape C de la procédure AC et à la suite des tests de type a. à c. Nous commençons par examiner la proposition ainsi spécifiée.

3.4.2.1 Un nouveau test à l'étape C

La proposition que nous envisageons maintenant est précisément la suivante : pour les hypothèses causales 1) qui ont passé les tests a. à c. et 2) pour lesquelles on ne peut pas rejeter l'hypothèse selon laquelle l'hypothèse de fidélité et la condition de Markov causale sont satisfaites, vérifier qu'elles font bien partie de l'ensemble de modèles représenté par la sortie de l'algorithme. Dans le cas où elle ne l'est pas, l'hypothèse causale discutée doit être rejetée ; dans le cas où elle l'est, elle s'en trouve corroborée.

A l'appui de cette proposition, on remarquera qu'il est bien possible de tester si la condition de Markov causale et l'hypothèse de fidélité seraient satisfaites dans le cas où un modèle causal donné serait adéquat. C'est ce que nous faisions valoir, déjà, à la fin du paragraphe 3.3.3.2. Ainsi, dans un contexte qui n'est plus a-théorique, nous ne sommes plus dans la situation inextricable que nous avons décrite dans le paragraphe 3.3.3.1.

Le test, bien sûr, ne serait pas complètement fiable : c'est une hypothèse statistique qu'il s'agirait d'invalider. Cela pourtant ne saurait compter contre la proposition que nous discutons. En effet, ainsi que nous l'avons vu dans le paragraphe 3.3.2, les difficultés associées au fait d'inférer des conclusions relatives à une population à partir de données dans un échantillon de cette population sont générales, non spécifiques d'un type d'inférence causale. En outre, elles sont incontournables ; en conséquence, le fait qu'elle ne donne pas de moyen de les contourner ne saurait compter contre notre proposition.

Maintenant, il nous faut examiner en quel sens un test C.d.[36] recourant

[36]Par « test C.d » nous désignons un test qui viendrait prendre place à l'étape C de l'inférence causale AC après les tests de type c.

aux algorithmes d'inférence RB peut être instructif. Or, à ce point, les choses deviennent moins favorables. En effet, nous savons maintenant que l'inférence causale AC repose sur une conception du rapport entre causalité et probabilités similaire à celle que définissent ensemble la condition de Markov causale et l'hypothèse de fidélité. Nous avons indiqué que cette conception est mobilisée au moment où les hypothèses sont spécifiées, à l'occasion des tests de type a. et à l'occasion des tests de type b. Dans ces conditions, une hypothèse qui aurait été spécifiée en A et ne serait pas rejetée à l'issue des tests de type a., b., c. ne peut pas manquer d'appartenir à l'ensemble des graphes que le résultat d'un algorithme RB représente. Aussi, un test dont l'issue dépendrait de ce qu'une hypothèse non rejetée après Cc. appartienne ou non au résultat des algorithmes RB n'instruirait pas.

A cette critique, il est possible de répondre en deux temps. Dans le premier, nous revenons sur le statut de la conception du rapport entre causalité et probabilités dans l'inférence AC. Plus précisément, nous revenons sur ceci, que nous rappelions dans le dernier paragraphe, que cette conception est mobilisée localement. Corrélativement, elle semble dispensable : elle n'intervient que dans des zones délimitables de l'inférence, et n'en constitue pas le principe. Dans ces conditions, on pourrait imaginer d'accompagner la proposition d'un test 3d. fondé sur les algorithmes RB de la préconisation de ne pas recourir à des hypothèses relatives au rapport entre la causalité et les probabilités *avant* le test. C'est ce qu'on fait si 1) on ne fait pas dépendre la formulation d'hypothèses causales (à l'étape A) de considérations probabilistes – mas seulement théoriques ; 2) on réalise le test RB non pas après, mais avant les tests Ca. à Cc.

Toutefois – et ce sera là le second temps de notre réponse – cela est sans compter que la façon dont les paramètres structurels sont estimés (ici à l'étape B) semble elle aussi engager une conception du rapport entre causalité et probabilités de type markovien. Plus précisément, le paramètre mesurant l'effet d'une variable C sur une variable E est estimé en tenant fixée la valeur des autres causes supposées de E. A proprement parler, il n'y a rien ici qui requiert que la condition de Markov causale soit satisfaite. Aussi n'avons-nous pas mentionné ce point dans le paragraphe 3.3.3.2. Toutefois ce mode d'estimation des paramètres ne fait sens que si on a l'idée que ses causes prises ensemble suffisent à expliquer les variations des valeurs d'une variable et ses corrélations avec les autres variables.

Dans ces conditions, prendre au sérieux l'injonction d'ignorer le rapport supposé entre la causalité et les probabilités jusqu'à la mise en oeuvre du test fondé sur les réseaux bayésiens, conduit à renoncer à la fois à l'interprétation causale des paramètres estimés à l'étape B. et aux tests a. à c. qui les concernent. En ce sens, la proposition d'intégrer les algorithmes

RB à l'étape C sans recourir auparavant à des hypothèses relatives au rapport entre causalité et probabilités implique tout bonnement de renoncer à l'inférence AC elle-même. L'injonction de prendre en compte dans le cadre AC la possibilité que la condition de Markov causale et l'hypothèse de fidélité soient violées n'est pas compatible avec le fait de recourir aux algorithmes RB si tard dans la procédure d'inférence causale.

3.4.2.2 Un test entre l'étape A et l'étape B

L'idée que nous envisageons maintenant est la suivante : utiliser les algorithmes RB pour tester les hypothèses causales dès après leur formulation. Plus précisément, la proposition n'a de sens que si les hypothèses causales sont formulées indépendamment de considérations probabilistes. En outre, le test ne peut être mené que pour les hypothèses acycliques et telles qu'on ne peut pas rejeter l'hypothèse selon laquelle elles satisfont la condition de Markov causale et l'hypothèse de fidélité. Pour de telles hypothèses, le test consisterait comme plus haut à vérifier que l'hypothèse causale envisagée appartient bien à l'ensemble de graphes orientés acycliques qui constituent la sortie d'un algorithme RB. Si ce n'est pas le cas, le modèle pourrait être rejeté dès l'étape A'. que nous envisageons ici.

Cette proposition présente les mêmes avantages que celle que nous envisagions dans le dernier paragraphe. Elle permet d'utiliser les algorithmes RB dans le cadre de l'inférence causale AC, et de les utiliser seulement quand les hypothèses corrélatives de la notion de réseau bayésien causal sont satisfaites. Mais, d'un autre côté, la présente proposition n'a pas les inconvénients qui nous ont arrêtés plus haut. D'un côté, le test est bien instructif et doit permettre de rejeter certaines hypothèses causales dès lors que la formulation de ces hypothèses (étape A.) ne repose pas sur des considérations probabilistes. D'un autre côté, il est compatible avec le fait de prendre en compte, dans le cadre AC, la possibilité que la condition de Markov causale et l'hypothèse de fidélité soient violées. Même, il contribue à cette prise en compte en impliquant qu'on recherche dès la formulation d'un modèle si ces hypothèses seraient satisfaites dans le cas où le modèle serait adéquat.

La proposition ne sera complète qu'une fois qu'on aura précisé ce qu'il convient de faire si le modèle spécifié est tel que la condition de Markov causale ou l'hypothèse de fidélité ne serait pas satisfaite dans le cas où il serait adéquat. A cette question, nous offrons deux réponses différentes, correspondant à la place différente que les deux hypothèses occupent dans l'inférence causale AC. Prendre en compte le fait que l'hypothèse de fidélité serait violée dans le cas où le modèle serait correct implique seulement de ne pas le rejeter à l'issue des tests Ca. si on ne peut pas rejeter l'hypothèse selon laquelle tous

les paramètres sont significativement différents de zéro. Le cas de la condition de Markov causale est différent, puisque nous avons vu qu'elle est sous-jacente dans nos méthodes d'estimation des paramètres. Prendre en compte une violation de la condition de Markov causale, c'est pratiquement renoncer à l'inférence AC même. Dans ces conditions, ne pas rejeter immédiatement un modèle tel que la condition de Markov causale serait violée dans le cas où il serait adéquat, requiert d'avoir de très bons arguments théoriques en faveur de cette hypothèse.

Reste, finalement, la question computationnelle. Elle ne peut pas être complètement résolue : les algorithmes RB étant ce qu'ils sont, leur complexité ne varie pas. A l'attention du lecteur légitimement inquiet, nous formulons néanmoins trois remarques. En premier lieu, nous avons pris soin de préciser que le test reposant sur un algorithme RB n'est pas mené pour chaque hypothèse causale. Il est mené seulement pour les hypothèses acycliques et telles qu'on ne peut pas rejeter l'hypothèse selon laquelle la condition de Markov causale et de l'hypothèse de fidélité sont satisfaites. En deuxième lieu, il convient de souligner que l'inférence causale RB n'a pas à être menée à nouveaux frais à chaque utilisation du test. En effet, le résultat d' un algorithme RB est inchangé par l'hypothèse qu'il y a à tester. En troisième lieu, enfin, nous noterons que si des hypothèses sont rejetées lors de l'étape A'. que nous envisageons, alors il n'est nul besoin de mener pour ces hypothèses les étapes B. et C. de la procédure décrite dans la sous-section 3.2.2.

3.5 Conclusion

Finalement, nous avons montré que les réseaux bayésiens ne renouvellent pas l'épistémologie de la causalité au sens où leurs partisans peuvent l'entendre. Il est vrai que, d'un point de vue logique, l'inférence causale RB se distingue de l'inférence menée grâce aux outils traditionnels de l'analyse de chemins par sa déductivité et que cette déductivité est corrélative de l'a-théoricité – ce que certains ont appelé « inductivité ». En revanche il est faux que ces caractéristiques logiques remarquables puissent profiter effectivement au projet d'inférence causale. L'argument nodal est ici le suivant : en raison même de son a-théoricité, l'inférence causale RB est telle qu'on doute toujours de la correction de ses résultats effectifs.

Toutefois, on peut définir un apport des réseaux bayésiens à l'épistémologie de la causalité générique. D'abord, ils mettent au premier plan une conception du rapport entre causalité et probabilités qui est sous-jacente à nos pratiques d'inférence causale, généralement inaperçue et dont nous sa-

vons qu'elle admet des contre-exemples. Pour cette raison et parce que les algorithmes RB sont des outils théoriques puissants, les réseaux bayésiens nous invitent plus généralement à redéfinir nos procédures d'inférence causale. Dans la dernière section, nous nous sommes essayé à définir une procédure d'inférence causale qui intègre les algorithmes RB et les analyses de la sous-section 3.3.3. A nos yeux, c'est sous la forme d'un test mené entre les étapes A et B que les réseaux bayésiens trouvent le plus raisonnablement leur place dans la procédure AC.

Chapitre 4

Condition de Markov causale et indéterminisme

Dans les deux derniers chapitres, nous avons mis au jour les modalités de l'inférence aux causes telle qu'elle est autorisée par les réseaux bayésiens. La question a été traitée d'abord du point de vue de l'analyse conceptuelle, puis du point de vue de la mtéhodologie de l'inférence aux causes génériques. De l'un et de l'autre points de vue, les principales caractéristiques de l'inférence causale fondée sur les réseaux bayésiens dépendent essentiellement de ce qu'elle ne donne des résultats corrects qu'à la condition que l'hypothèse d'acyclicité, la condition de Markov causale et l'hypothèse de fidélité soient satisfaites.

L'enquête menée dans le chapitre qui commence est orthogonale des enquêtes menées dans les chapitres 2 et 3, puisqu'elle porte précisément sur la question de savoir quand ces hypothèses sont satisfaites. Ainsi, il ne s'agit plus d'explorer les conséquences épistémologiques de ce que l'inférence causale RB repose sur ces hypothèses, mais d'essayer de déterminer quand elles le sont. Le chapitre qui commence prend donc place dans le débat sur l'extension du domaine au sein duquel les hypothèses corrélatives des méthodes d'inférence causale qui nous intéressent sont vraies. Plus précisément, il prend place dans le champ de la discussion de la condition de Markov causale. Nous avons indiqué déjà qu'elle est celle de nos trois hypothèses qui a fait couler le plus d'encre.

Cette discussion consiste en partie dans la mise en évidence de classes de systèmes qui satisfont prouvablement la condition. Or, à cet endroit, les systèmes indéterministes revêtent un caractère crucial. D'un côté, en effet, les résultats de validité de la condition de Markov concernent des classes de systèmes déterministes et l'indéterminisme se présente comme une source difficilement tarissable d'exemples solides de violation de la condition de Markov

causale.[1] De l'autre côté, les théories probabilistes de la causalité se légitiment spécifiquement par l'existence d'effets que leurs causes ne suffisent pas à produire – ce que, dans le contexte présent, on appellera « indéterminisme».

Dans Steel (2005), Daniel Steel prétend étendre du cas déterministe au cas indéterministe le résultat selon lequel les systèmes dont les variables exogènes sont conjointement indépendantes[2] satisfont la condition de Markov causale. On aura compris que la thèse de Steel, si elle est fondée, contribue de manière significative au débat relatif à la condition de Markov causale en particulier et, par extension, à l'inférence aux causes génériques fondée sur les réseaux bayésiens.

Le chapitre qui commence vise précisément à déterminer si la thèse de Steel est fondée et, plus généralement, si et en quel sens les systèmes déterministes sont plus susceptibles que les systèmes indéterministes de satisfaire la condition de Markov causale. L'analyse que nous proposons compte quatre temps :

1. dans la première section, nous présentons les résultats classiques de satisfaction de la condition de Markov causale par certaines classes de systèmes déterministes ;
2. dans la deuxième section, nous présentons le résultat établi dans Steel (2005) ;
3. dans la troisième section, nous critiquons la présentation que Steel donne du résultat qu'il établit. Plus précisément, nous montrons que ce n'est qu'au prix d'une définition inhabituelle de certains termes que le résultat établi par Steel est bien celui qu'il annonce ;
4. dans la quatrième section, nous mettons au jour une innovation méthodologique suggérée par Steel (2005), et prolongeons la suggestion jusqu'à établir que le déterminisme est bien, finalement, plus favorable que l'indéterminisme pour la condition de Markov causale.

4.1 Déterminisme et condition de Markov causale

Dans la section qui commence, nous présentons les résultats classiques de satisfaction de la condition de Markov causale par certains systèmes déterministes. Cela n'est possible qu'après avoir expliqué plus précisément

[1] Ainsi, nous avons indiqué dans le paragraphe 1.3.2.1 comment l'indéterminisme engendre des contre-exemples à la troisième composante de l'hypothèse de Markov causale.

[2] Nous expliquerons dès que possible (c'est-à-dire dans la première section du présent chapitre) ce qu'il faut entendre par là.

de quoi il est question ici, c'est-à-dire après avoir défini et discuté la notion de déterminisme à laquelle il est fait référence. C'est ce que nous faisons dans la première sous-section.

4.1.1 Déterminisme

Dans le champ qui nous intéresse ici, il n'est pas clair quels objets sont susceptibles d'être déterministes ou indéterministes. Ainsi Spirtes, Glymouret Scheines parlent successivement de : distribution de probabilités déterministe[3], graphe déterministe[4], réseau bayésien déterministe[5], système déterministe[6], structure causale déterministe[7], dispositif (*device*) déterministe[8], relation déterministe entre deux ensembles de variables[9]. La plupart de ces notions ne sont pas définies clairement, pas plus d'ailleurs que les rapports qu'elles entretiennent. En outre, aucune de ces notions n'est classique : c'est habituellement du monde qu'on se demande s'il est ou non déterministe.[10] Dans ces conditions, la tâche qui nous incombe consiste d'abord à proposer une définition du déterminisme qui permette de rendre compte de ses usages dans le débat sur la condition de Markov causale, et ensuite à indiquer comment la définition qu'on aura proposée s'articule d'une part avec les notions principales parmi celles qui apparaissent dans l'énumération qui précède et d'autre part avec le concept plus classique de déterminisme du monde.

Ensemble de variables déterministe. Un examen des usages qui sont faits de la notion de déterminisme dans le débat sur la condition de Markov causale révèle que, dans ce contexte, ce sont essentiellement des ensembles de variables qui sont susceptibles de déterminisme. Plus précisément, il s'agit d'ensembles de variables sur lesquels le graphe causal[11] est acyclique. Puisque

[3]Spirtes et al. (1993) p.15.

[4]Spirtes et al. (1993) p. 15.

[5]Spirtes et al. (1993) p. 16. En fait c'est de réseau bayésien « pseudo-indéterministe » qu'il est question en ce point. Seulement nous ne voyons pas pour les objets du déterminisme devraient différer de ceux du pseudo-indéterminisme. Nous considérons donc que s'il existe des réseaux bayésiens pseudo-indéterministes, alors il existe des réseaux bayésiens déterministes.

[6]Spirtes et al. (1993) pp. 16 et 32.

[7]Spirtes et al. (1993), p. 25.

[8]Spirtes et al. (1993) p. 25.

[9]Spirtes et al. (1993) p. 53.

[10]A titre d'illustration de cette thèse, on peut parcourir Earman (1986).

[11]Par « graphe causal» nous entendons désormais le graphe défini sur l'ensemble de variables qu'on considère et tel qu'il existe une flèche de V_i à V_j exactement quand V_i est

c'est la condition de Markov causale et non la troisième composante de l'hypothèse de représentation qui est examinée dans ce chapitre, nous ne remettons pas en cause l'hypothèse d'acyclicité. Aussi, tout ce qui est dit dans la suite de ce chapitre concerne les seuls ensembles de variables sur lesquels le graphe causal est acyclique – et nous ne prendrons donc plus la peine de le préciser.

En vue de définir ce qu'est un ensemble de variables déterministe, il convient d'introduire les notions de variables endogènes et exogènes :

Définition 4.1 (Variables endogènes et exogènes) *Soit* $\boldsymbol{V}$ *un ensemble de variables.*
Sont endogènes *dans* $\boldsymbol{V}$ *les variables de* $\boldsymbol{V}$ *qui ont au moins une cause directe dans* $\boldsymbol{V}$*.*
Sont exogènes *dans* $\boldsymbol{V}$ *les variables de* $\boldsymbol{V}$ *qui n'ont pas de cause directe dans* $\boldsymbol{V}$*.*

On notera que, de même que la notion de cause directe, l'endogénéité et l'exogénéité d'une variable donnée sont relatives à l'ensemble de variables $\mathbf{V}$ qu'on considère.

A partir de ces premières définitions, on peut en venir à celle qui nous intéresse plus directement :

Définition 4.2 (Ensemble de variables déterministe) *Un ensemble de variables* $\boldsymbol{V}$ *est* déterministe *si la valeur des variables endogènes de* $\boldsymbol{V}$ *est déterminée fonctionnellement par celle de leurs causes directes dans* $\boldsymbol{V}$*. Dans le cas contraire, il est indéterministe.*

Cette définition demande à être explicitée sur deux points. En premier lieu, précisons qu'il est question dans cet énoncé de la valeur que les variables de $\mathbf{V}$ prennent pour un individu. Plus précisément encore, cet individu appartient à la population relativement à laquelle existent les relations de cause à effet auxquelles il est fait référence. En second lieu, insistons sur ceci que cette définition générale du déterminisme n'impose aucune restriction sur la forme de la fonction de détermination de la valeur d'une variable d'un ensemble déterministe par celles de ses causes directes dans cet ensemble.

A partir de la définition 4.2, on peut rendre compte du rapport qui existe entre le déterminisme et :

– les distributions de probabilités : toute distribution de probabilités sur un ensemble de variables $\mathbf{V}$ déterministe et sur lequel le graphe causal

une cause directe de V_j. En effet, nous avons montré dans la section 1.3 que les artifices discutés dans le chapitre précédent n'ont pas leur place dans le contexte (d'inférence aux causes génériques) qui nous intéresse ici.

est acyclique est complètement définie par la distribution de probabilités marginale sur l'ensemble **VE** des variables exogènes de **V**. En effet, toute valeur de **V** est soit telle que la valeur des variables endogènes est compatible (au sens défini dans l'appendice au chapitre 1) avec celle de **VE**, soit telle que la valeur des variables endogènes n'est pas compatible celle de **VE**. Dans le premier cas, sa probabilité est celle de la valeur de **VE** ; dans le second cas, elle est nulle ;

– les graphes orientés acycliques : le graphe causal sur un ensemble de variables déterministe est tel que la valeur des variables qui ont des parents dans ce graphe est déterminée par la valeur de ces parents.

Il reste toutefois ceci que le déterminisme ne se présente pas usuellement et d'abord comme une propriété d'ensembles de variables, mais comme une propriété sinon peut-être immédiatement du monde, du moins des systèmes réels que ces ensembles de variables contribuent à représenter. Il convient dès lors de se demander comment la notion de déterminisme que nous avons définie s'articulent à des propriétés de systèmes réels.

Déterminisme et systèmes réels. Considérons un ensemble de variables **V**. Chacune de ces variables, nous l'avons vu dans le chapitre 1 (paragraphe 1.2.2.1), représente une famille de propriétés. Nous avons vu également que les relations de cause à effet entre variables découlent des relations de cause à effet entre propriétés : dire que la variable V_i cause la variable V_j, c'est dire au moins qu'une propriété de la famille représentée par V_i cause une propriété de la famille représentée par V_j. Par ailleurs, dire que la propriété P_1 cause la propriété P_2, c'est dire qu'il existe un mécanisme en vertu duquel le fait qu'un individu possède P_1 a une influence sur le fait qu'il possède P_2.[12] Finalement on peut caractériser un système réel comme un ensemble d'objets régi par de tels mécanismes, qui est relativement isolé et qu'il est pertinent d'analyser pour lui-même. Bien sûr, nous ne prétendons pas que ceci soit une définition satisfaisante, mais plutôt une explicitation de ce qu'on entend couramment par « système », qui suffit à donner sens aux analyses du présent chapitre. Ainsi, cette explicitation ne vise pas à rendre compte de l'ensemble des usages du terme, mais seulement de l'usage qu'on en fait quand on parle par ailleurs de causalité entre variables.

D'un système réel caractérisé ainsi que nous venons de le faire, on pourra dire qu'il est déterministe si la situation d'un individu relativement à certaines propriétés dont l'instanciation dépend du système – qu'on pourrait

[12] Aussi vague soit-il, le terme « influence» est choisi à dessein et d'ailleurs en raison même de son indétermination. Pour des raisons similaires, le terme « mécanisme» n'est pas spécifié plus avant.

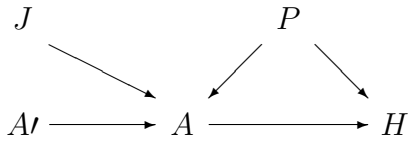
appeler « propriétés antécédentes» – détermine sa situation relativement à toutes. Cette caractérisation est inspirée de la définition 4.2. Elle lui correspond d'ailleurs exactement, s'il est vrai qu'un système réel est déterministe exactement quand est déterministe l'ensemble des variables représentant les familles de propriétés observables dont l'instanciation dépend du système.

Cette analyse, toutefois, n'est correcte qu'à la condition que toutes les propriétés observables dont l'instanciation dépend du système soient représentées par les variables de l'ensemble choisi pour représenter ce système. Pour comprendre ce qui est en jeu ici, il convient de considérer un exemple. Celui que nous proposons est librement inspiré d'un exemple souvent discuté par Pearl.[13] Soit donc, à titre d'illustration, le système d'arrosage d'une pelouse. La pelouse peut être arrosée de deux façons différentes : soit par un jardinier au moyen d'un arrosoir, soit par un arroseur électrique. Celui-ci peut être déclenché manuellement ou fonctionner en mode automatique. Dans ce dernier mode, il arrose la pelouse à une heure fixée et à condition qu'il ne pleuve pas. Par ailleurs, l'arrosage et la pluie sont les deux seules raisons pour lesquelles la pelouse peut être humide. Il nous semble clair qu'on peut considérer ce système comme déterministe : rien ne s'y passe qui ne soit pas déterminé par ce qui le cause.

Considérons maintenant l'ensemble de variables $\mathbf{V} = \{A', A, H\}$ tel que A' est la variable représentant l'état de l'arroseur automatique et A et H représentent respectivement si la pelouse est arrosée et si la pelouse est humide. A' est susceptible de prendre les trois valeurs : « éteint», « mode manuel» et « mode automatique» ; A et H sont binaires, prenant la valeur 1 si ce qu'elles représentent est le cas et la valeur 0 sinon. Chacune de ces trois variables représente une famille de propriétés observables dont l'instanciation dépend du système ; en ce sens, l'ensemble $\mathbf{V}$ qu'elles constituent représente le système réel considéré. Pourtant, $\mathbf{V}$ n'est pas déterministe. En particulier, A a A' pour seule cause directe dans $\mathbf{V}$, mais peut prendre des valeurs différentes pour la même valeur « éteint» de A' : pour la valeur « éteint» de A', A prend la valeur 1 quand le jardinier arrose la pelouse et peut prendre la valeur 0 sinon.

Mais considérons maintenant $\mathbf{V'} = (\mathbf{V} \cup \{J, P\})$ où J et P représentent respectivement si le jardinier a arrosé la pelouse et s'il pleut. $\mathbf{V'}$ aussi représente le système réel auquel nous nous intéressons. Or, le graphe causal sur cet ensemble est :

[13] Il l'est en particulier dans Pearl (2000) p.15.



et on vérifie aisément que la valeur des variables de ce graphe qui ont des parents dans ce graphe est déterminée par la valeur de ces parents. Par conséquent, l'ensemble de variables **V'** représentant le système qu'on considère est déterministe. De façon plus générale, si un système S est déterministe, il est faux que tout ensemble de variables représentant S est déterministe, mais il est vrai que l'est un ensemble de variables $\mathbf{V}_S$ représentant toutes les propriétés observables pertinentes dont l'instanciation dépend du système. Par ailleurs, un système peut être indéterministe parce qu'il existe une fourche interactive composée de variables d'un ensemble qui suffit à représenter toutes les propriétés observables dont l'instanciation dépend du système. Corrélativement, une classe importante de violations de la troisième composante de l'hypothèse de Markov causale se caractérise comme indéterministe.

Maintenant que nous avons mis en évidence le rapport qui existe entre déterminisme d'un ensemble de variables et déterminisme d'un système réel, nous pouvons en venir au rapport que ce dont nous parlons entretient avec le déterminisme comme thèse sur le monde. En effet, il nous semble que c'est du déterminisme d'un système réel que se rapproche le plus ce que serait le déterminisme du monde. D'un côté, en effet, nous avons vu qu'un système réel est déterministe si la situation d'un individu relativement aux propriétés antécédentes détermine sa situation relativement à toutes les propriétés dont l'instanciation dépend du système réel. Ainsi, deux individus qui sont dans la même situation relativement aux propriétés antécédentes sont dans la même situation relativement à toutes les propriétés. De l'autre côté, une définition simple mais déjà relativement robuste qu'on peut donner du déterminisme comme propriété éventuelle du monde est la suivante : le monde est déterministe si tout monde physiquement possible qui coïncide avec lui en un instant t, coïncide avec lui en tous les instants postérieurs à t[14]. On aperçoit alors les points saillants d'une analogie entre le déterminisme d'un système réel et le caractère d'un univers de mondes physiquement possibles déterministes. Surtout, on entrevoit la différence qui existe entre le déterminisme tel qu'il nous intéresse ici et le déterminisme au sens habi-

[14]Earman (1986) p.13. La définition que nous adoptons est celle du « déterminisme laplacien relativement au futur» (*futuristically Laplacian determinism*), qui nous paraît à la fois celui qui correspond le mieux à nos intuitions relatives au déterminisme et celui pour lequel le rapport avec le déterminisme d'un système réel apparaît le plus clairement.

tuel. Nous ne dressons pas une typologie de ce qui les oppose, mais nous contentons d'en inférer que parler de système réel déterministe – et donc d'ensemble de variables déterministe – au sens où nous le faisons ici ne suppose pas d'avoir répondu positivement à la question de savoir si le monde est, ou non, déterministe.

4.1.2 Résultat classique

En vue d'énoncer le résultat classique de satisfaction de la condition de Markov causale par certains systèmes déterministes, il nous faut définir la notion probabiliste d'indépendance conjointe :

Définition 4.3 (Indépendance conjointe) *Les variables d'un ensemble* $\boldsymbol{V}$ *sont* conjointement indépendantes *pour* p *si pour tout couple* ($\boldsymbol{W}$, $\boldsymbol{X}$) *de sous-ensembles de* $\boldsymbol{V}$ *non vides et d'intersection nulle,* $\boldsymbol{W}$ *est (absolument) indépendant pour* p *de* $\boldsymbol{X}$.

Le résultat qui nous intéresse ici peut alors s'énoncer comme suit :

Théorème 4.1 (Ensembles de variables déterministes et CMC) *Les ensembles de variables déterministes dont les variables exogènes sont conjointement indépendantes satisfont la condition de Markov causale.*

Sous la double convention que les variables exogènes d'un système S sont les variables exogènes de l'ensemble $\mathbf{V}_S$ qui représentent toutes les propriétés observables dont l'instanciation dépend de S et que S satisfait la condition de Markov causale si et seulement si $\mathbf{V}_S$ la satisfait, le théorème 4.1 reçoit sa formulation usuelle : les systèmes déterministes dont les variables exogènes sont conjointement indépendantes satisfont la condition de Markov causale.

Si ce résultat est régulièrement mentionné dans la littérature, nous n'avons pas connaissance d'une démonstration complète du théorème 4.1. Aussi en proposons-nous une[15] :

Preuve : Soit $\mathbf{V}$ un ensemble de variables déterministe, G le graphe causal sur $\mathbf{V}$ et p la distribution de probabilités sur $\mathbf{V}$. Supposons que les variables de $\mathbf{V}$ sont conjointement indépendantes pour p.
$\mathbf{V}$ satisfait la condition de Markov causale si et seulement si toute variable de $\mathbf{V}$ est indépendante pour p de ses non-descendants dans G relativement à l'ensemble de ses parents dans $\mathbf{G}$. Soit donc une variable V de $\mathbf{V}$; on va montrer qu'elle est effectivement indépendante pour p de ses non-descendants dans G relativement à

[15] On trouve une démonstration similaire dans Steel (2005) pp. 22–23, mais pour un résultat légèrement différent et que nous souhaitons distinguer soigneusement de celui qui nous intéresse dans la présente sous-section.

l'ensemble de ses parents dans G.
V peut être soit endogène, soit exogène dans $\mathbf{V}$.

- Si V est endogène dans $\mathbf{V}$, sa valeur est déterminée fonctionnellement par celle de ses parents dans G. Donc V est indépendante pour p et relativement à l'ensemble de ses parents dans G de toutes les variables de $\mathbf{V}$ qui ne sont pas ses descendants dans G.
- Si V est exogène dans $\mathbf{V}$, elle n'a pas de parent dans G. Ce qu'il y a à montrer est donc que V est indépendante absolument de ses non-descendants dans G. Soit W une variable de $\mathbf{V}$ qui n'est pas un descendant de V dans G.
  - Si W est elle aussi exogène dans $\mathbf{V}$, l'indépendance de V et de W découle de l'hypothèse d'indépendance conjointe des variables exogènes de $\mathbf{V}$.
  - Si W est endogène dans $\mathbf{V}$, par l'hypothèse selon laquelle $\mathbf{V}$ est déterministe et parce que G est acyclique, sa valeur est fonctionnellement déterminée par celle de ces ancêtres dans G qui sont exogènes relativement à $\mathbf{V}$. Comme W n'est pas un descendant de V, V n'appartient pas à cet ensemble. L'indépendance conjointe des variables exogènes de $\mathbf{V}$ implique donc que V est indépendante de l'ensemble de ces variables, et donc de W dont il détermine la valeur.

  Dans les deux cas, V est indépendante de W**.**

Ainsi, que V soit endogène ou exogène, elle est indépendante de ses non-descendants dans G relativement à l'ensemble de ses parents dans G**.**

Maintenant que nous avons présenté l'élément principal du contexte théorique auquel Steel (2005) se réfère, nous pouvons en venir à l'article lui-même.

4.2 Le résultat de Steel

Dans Steel (2005), l'auteur prétend étendre le résultat de satisfaction de la condition de Markov causale par les systèmes dont les variables exogènes sont conjointement indépendantes du cas déterministe que nous venons de détailler, au cas indéterministe. Steel prétend donc montrer que la condition de Markov causale est satisfaite par tout système – déterministe ou non – dont les variables exogènes sont conjointement indépendantes. Dans la section qui commence, nous présentons le résultat de Steel. Plus précisément, et puisque tous les termes qui apparaissent dans l'énoncé du résultat de Steel ont été expliqués dans la section précédente, nous nous attachons à rendre compte de la façon dont Steel établit ce résultat. Pour cela, nous procédons en trois temps : d'abord nous présentons le cadre théorique adopté par Steel, ensuite nous détaillons ce que Steel montre dans ce cadre, enfin nous aurons à préciser comment le résultat établi s'articule avec la notion d'indéterminisme.

Avant de commencer, notons qu'il apparaîtra rapidement que le cadre théorique adopté par Steel diffère sensiblement du cadre classique que nous

avons introduit dans la section précédente. Si cette différence semble appeler une analyse minutieuse et une discussion serrée, ce n'est que dans la prochaine section que celles-ci seront menées. Dans la présente section, nous concentrons notre attention sur le seul résultat établi dans Steel (2005) et nous attachons à le rendre intelligible.

4.2.1 Modèles fonctionnels causaux

Le cadre théorique choisi par Steel pour établir le résultat qui fait l'objet de Steel (2005) est constitué de « modèles fonctionnels causaux». En vue de comprendre ce qu'est un modèle fonctionnel causal, il convient de définir d'abord les modèles fonctionnels :

Définition 4.4 (Modèles fonctionnels (Steel, 2005)) *Soit* $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ *et* $\boldsymbol{U} = \{U_1, U_2, \ldots, U_k\}$ *deux ensembles disjoints de variables.*
Un modèle fonctionnel sur (**X**, **U**) *est une paire (**E**, p) où :*

- ***E** est un ensemble de n équations tel que chaque variable X_i apparaît comme une fonction f_i des variables d'un sous-ensemble non vide de $((\boldsymbol{X} \cup \boldsymbol{U}) \setminus \{X_i\})$;*
- *p est une distribution de probabilités sur **U**.*

A titre d'illustration, on peut définir un modèle fonctionnel M_E sur $(\{X_1, X_2, X_3), \{U_1, U_2\})$ par les équations :

$$X_1 = f_1(U_1, U_2)$$
$$X_2 = f_2(X_1, U_2)$$
$$X_3 = f_3(X_2)$$

et une distribution de probabilités sur $\{U_1, U_2\}$. Cet exemple met en lumière ceci que la définition 4.4 des modèles fonctionnels n'implique pas l'existence d'une correspondance bi-univoque entre les ensembles **X** et **U** qui composent le couple sur lequel un modèle fonctionnel est défini.

Venons-en maintenant aux modèles fonctionnels causaux. La définition proposée par Steel[16] est la suivante :

Définition 4.5 (Modèles fonctionnels causaux (Steel, 2005)) *Un modèle fonctionnel (**E**, p) sur (**X**, **U**) est* causal *si :*

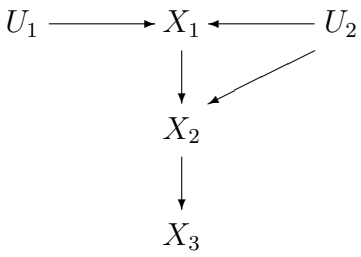
1. *les équations de **E** sont des généralisations causales ;*

[16] Steel (2005) p.9.

2. *toute variable X_i de $\boldsymbol{X}$ est telle que l'ensemble $\boldsymbol{DC}(X_i)$ de ses causes directes dans ($\boldsymbol{X} \cup \boldsymbol{U}$) est inclus dans l'ensemble des variables dont elle est une fonction dans $\boldsymbol{E}$.*

Ce qu'est l'apport exact de sa première clause à cette définition ne nous apparaît pas clairement. Nous retenons néanmoins cette clause afin de ne pas risquer de dénaturer la pensée de Steel.

En vue d'énoncer le résultat établi dans Steel (2005) et d'en exposer la preuve, trois séries de remarques relatives aux définitions 4.4 et 4.5 sont encore nécessaires. En premier lieu, et ainsi que le note Steel lui-même, il existe une « correspondance immédiate entre les modèles fonctionnels et les graphes orientés»[17]. Plus précisément, pour tout modèle fonctionnel $M = (\mathbf{E},p)$ sur ($\mathbf{X}$, $\mathbf{U}$), il existe un unique graphe orienté G_M tel que les parents d'une variable dans G_M sont exactement les variables dont elle dépend fonctionnellement selon $\mathbf{E}$.[18] Concernant le modèle fonctionnel M_E que nous avons partiellement défini plus haut, le graphe est le suivant :



Ce graphe orienté est acyclique. Plus généralement, on supposera que G_M est acyclique pour tout modèle fonctionnel M, causal ou non, que nous discuterons. Cette hypothèse correspond (en un sens qui sera rendu complètement précis plus bas) à la décision que nous avons prise au début de la dernière section, de ne considérer que des ensembles de variables sur lesquels les graphes causaux sont acycliques.

De ce que le graphe qui correspond à un modèle fonctionnel $M = (\mathbf{E}, p)$ sur ($\mathbf{X}$, $\mathbf{U}$) est acyclique, il découle – et ce sera là notre deuxième remarque – que les équations de $\mathbf{E}$ peuvent être réécrites de telle façon que les variables de $\mathbf{X}$ apparaissent comme des fonctions des seules variables de $\mathbf{U}$.[19] Dans ce cas, p détermine univoquement son extension p' à ($\mathbf{X} \cup \mathbf{U}$). En ne considérant que des modèles fonctionnels auxquels correspondent des graphes acycliques, on s'assure donc en particulier que les variables de $\mathbf{X}$ sont des fonctions des

[17]Steel (2005) p. 7.

[18]On notera qu'il découle de cette définition que les variables de $\mathbf{U}$ n'ont pas de parent dans G_M.

[19]Sur ce point, voir Steel (2005) p.8.

seules variables de **U** et que p détermine une distribution de probabilités sur ($\mathbf{X} \cup \mathbf{U}$) – et donc marginalement sur **X** (au sens de la définition 1.13).

En troisième et dernier lieu on notera que le graphe G_M qui correspond à un modèle fonctionnel *causal* M sur (**X**,**U**) est un sur-graphe du graphe causal GC_M sur $\mathbf{DC}_M = \mathbf{X} \cup (\bigcup_{X \in \mathbf{X}} \mathbf{DC}(X))$. Notons que Steel suppose qu'aucune variable de $((\mathbf{X} \cup \mathbf{U}) \setminus \mathbf{DC}_M)$ n'est un parent commun à plusieurs variables de G_M. Nous ne remettons pas en cause cette hypothèse. La justification proposée par Steel est la suivante : puisqu'aucune variable de $((\mathbf{X} \cup \mathbf{U}) \setminus \mathbf{DC}_M)$ n'est cause directe d'une variable de **X**, les caractéristiques des variables de $((\mathbf{X} \cup \mathbf{U}) \setminus \mathbf{DC}_M)$ « sont de pures conventions et non des hypothèses substantielles»[20]. Pour rendre cette justification convaincante, nous l'explicitons dans les termes suivants : toute dépendance probabiliste entre deux variables X et Y qui serait impliquée par un parent commun dans ($\mathbf{X} \cup \mathbf{U}) \setminus \mathbf{DC}_M$) – disons : U – peut être représentée comme une dépendance probabiliste entre deux variables U_X et U_Y qu'on substitue à U dans **U**, qui prennent les mêmes valeurs qu'elle et qui sont telles que U_X est parent de la seule variable X et U_Y est parent de la seule variable Y.[21] Ces remarques achèvent notre présentation du cadre théorique adopté par Steel, et nous pouvons maintenant en venir au résultat qu'il établit.

4.2.2 Ce que Steel montre (et comment il le montre)

Le résultat établi dans Steel (2005) peut être énoncé de la façon suivante :

Théorème 4.2 (Steel, 2005) *Soit $M = (\boldsymbol{E},p)$ un modèle fonctionnel causal sur $(\boldsymbol{X},\boldsymbol{U})$.*
*Si les variables de **U** sont conjointement indépendantes pour p, alors $\boldsymbol{DC}_M$ satisfait la condition de Markov causale.*

La preuve proposée par Steel[22], et dont la structure est décrite d'abord par Pearl pour un résultat un peu différent[23], s'explicite de la façon suivante :

Preuve : Soit M = (**E**,p) un modèle fonctionnel causal sur (**X**,**U**), G_M le graphe orienté (acyclique) qui correspond à M et p' la distribution de probabilités sur ($\mathbf{X} \cup \mathbf{U}$) qui est déterminée p. On suppose que les variables de **U** sont conjointement indépendantes et on montre :

1. que le graphe G_M défini sur ($\mathbf{X} \cup \mathbf{U}$) représente p'. Cela se démontre selon des voies similaires à celles que nous avons empruntées pour démontrer le

[20] Steel (2005) p. 14.
[21] La généralisation au cas où une même variable U est un parent commun à n variables de **X** est immédiate.
[22] Steel (2005) pp. 22–23.
[23] Pearl (2000) p. 30.

théorème 4.1, le rôle que joue le déterminisme dans cette preuve-ci étant joué dans cette preuve-là par la détermination fonctionnelle ;

2. que les indépendances qui correspondent à la satisfaction de la condition de Markov causale par $\mathbf{DC}_M$ s'établissent alors par d-séparation dans le réseau bayésien (G_M, p'). Plus précisément, il s'agit de montrer que toute variable de $\mathbf{DC}_M$ est d-séparée dans G_M de ses non-descendants dans GC_M par l'ensemble de ses parents dans GC_M. Aussi, considérons une variable V quelconque de $\mathbf{DC}_M$.
   - Si V n'a pas de parent dans GC_M, il y a à montrer que, dans G_M, V est d-séparé (par l'ensemble vide) de ses non-descendants dans GC_M. Soit W un non-descendant de V dans GC_M. Comme V n'a pas de parent dans GC_M, W n'est pas un ancêtre de V dans GC_M, ni ne partage un ancêtre avec V dans GC_M. Il en découle que tous les chemins de G_M entre V et W contiennent une fourche inversée, et sont donc d-séparés ;
   - Si V a des parents dans GC_M, il y a à montrer que dans G_M l'ensemble de ces parents d-sépare V de ses non-descendants dans GC_M. Pour cela, considérons W non-descendant de V dans GC_M. Quatre cas sont possibles :
     (a) W est un ancêtre de V dans GC_M. Dans ce cas, W est un ancêtre de V dans G_M. En effet, rappelons que G_M est un sur-graphe de GC_M ;
     (b) V et W ont un parent commun dans GC_M. Dans ce cas, V et W ont ce parent en commun dans G_M par l'hypothèse selon laquelle les variables de $((\mathbf{X} \cup \mathbf{U}) \setminus \mathbf{DC}_M)$ ne sont pas des parents communs de plusieurs variables de $\mathbf{DC}_M$;
     (c) V et W ont dans GC_M un ancêtre en commun qui n'est pas un parent de V. Dans ce cas, V et W ont cet ancêtre en commun dans G_M et les variables qui sont entre lui et V dans GC_M sont entre lui et V dans G_M ;
     (d) V et W ont un descendant commun dans GC_M. Dans ce cas, V et W ont ce descendant en commun dans G_M.

     Dans tous les cas, V est d-séparé dans G_M de W par l'ensemble de ses parents dans GC_M.

   Par le théorème 1.1, il en découle que V est indépendante de ses non-descendants dans GC_M relativement à l'ensemble de ses parents dans GC_M : $\mathbf{DC}_M$ satisfait la condition de Markov causale**.**

Dans la sous-section qui s'achève ici, nous avons énoncé le théorème 4.2 de Steel et nous en avons donné une preuve. Un point, cependant, reste obscur : en quoi ce résultat peut-il être considéré comme une extension au cas indéterministe du résultat classique 4.1 ? En des termes plus généraux, la question se pose du rapport entre le résultat de Steel et la no-

tion d'indéterminisme. La dernière sous-section de la présente section est consacrée à essayer de lui apporter une réponse.

4.2.3 Le théorème 4.2 et l'indéterminisme

Le rapport entre le résultat établi par Steel et l'indéterminisme consiste tout entier dans ceci que les modèles fonctionnels causaux sur lesquels le résultat porte peuvent représenter des systèmes réels indéterministes aussi bien que déterministes. Pour le montrer, Steel prend un exemple :

> Imaginez une voiture d'un type spécial, la voiture quantique. L'allumage de la voiture quantique dépend d'un processus fondamentalement indéterministe : quand on tourne la clef, la probabilité que la voiture démarre est irréductible et vaut 0,85.[24]

En reprenant les notations utilisées par Steel, les propriétés observables dont l'instanciation dépend du système que constituent les mécanismes à l'oeuvre dans le démarrage de la voiture quantique sont représentées par les variables :

- X_1 qui prend la valeur 1 si la clef est tournée et la valeur 0 sinon ;
- X_2 qui prend la valeur 1 si la voiture démarre et la valeur 0 sinon.

L'ensemble de variables $\{X_1, X_2\}$ n'est pas déterministe au sens où nous avons défini ce terme dans la première section. La raison en est le caractère probabiliste de l'action de X_1 sur X_2 ou, en d'autres termes, le fait que X_1 est une cause indéterministe de X_2. Il en découle que le système considéré par Steel est bien indéterministe au sens lui aussi défini dans la première section.

Une fois introduit l'exemple de la voiture quantique, Steel explique comment le système indéterministe que constituent les mécanismes à l'oeuvre dans le démarrage de cette voiture peut être représenté au moyen d'un modèle fonctionnel causal. Ce modèle, qu'on notera M, est défini sur un ensemble de variables $(\{X_1, X_2\}, \{U_1, U_2\})$. X_1 et X_2 sont définies de la façon que nous venons de décrire. Les définitions de U_1 et U_2 sont moins claires :

- U_1 est « un parent exogène de X_1»[25]. Par là, nous comprenons qu'il s'agit d'une variable représentant des propriétés dont l'instanciation cause la rotation de la clef. Nous ne comprenons pas pourquoi ces causes n'ont pas été mentionnées dans la description des conditions de démarrage de la voiture quantique, mais nous admettons avec Steel leur existence ;
- U_2 est une variable binaire de valeur possible 0 et 1 dont il est seulement spécifié que ces valeurs ont pour probabilités respectives 0,15 et 0,85.

[24] Steel (2005) p.13.
[25] Steel (2005) p. 13.

M se compose alors des équations :

$$X_1 = f_1(U_1)$$
$$X_2 = U_2 \times X_1$$

et de la distribution de probabilités p sur $\{U_1, U_2)$ qui donne aux valeurs de U_1 leur probabilité physique objective et aux valeurs de U_2 les probabilités indiquées ci-dessus (respectivement 0,15 et 0,85).

La façon dont la probabilité marginale sur U_2 est définie indique clairement que U_2 représente le caractère probabiliste de l'action de X_1 sur X_2. De façon générale, l'exemple proposé par Steel donne à voir comment tout système indéterministe peut être représenté au moyen d'un modèle fonctionnel causal : il suffit pour cela de représenter au moyen de variables de **U** le caractère probabiliste de l'action des causes indéterministes sur leurs effets. De ce que tout système – déterministe ou non – peut être représenté au moyen d'un modèle fonctionnel causal, Steel infère que l'énoncé du théorème 4.2 revient à ceci que la condition de Markov causale est satisfaite par tout système dont les variables exogènes sont conjointement indépendantes – qu'il soit déterministe ou non. Dès lors, le résultat 4.2 est considéré comme une généralisation au cas indéterministe du résultat classique 4.1, qui concerne le seul cas déterministe.

Dans la section qui s'achève, nous avons défini sans le discuter le cadre théorique adopté par Steel et présenté son résultat. De ce résultat, nous avons en particulier donné une preuve. Cette preuve est celle que propose Steel ; nous l'avons seulement explicitée sur quelques points. Il en découle que nous ne remettons en cause ni la correction de la preuve de Steel, ni la vérité de sa conclusion. En revanche, nous nous apprêtons à critiquer l'inférence de ce que tout système peut être représenté par un modèle fonctionnel causal, à la thèse selon laquelle le résultat 4.2 est une généralisation du résultat classique 4.1. Plus explicitement, nous soutenons que Steel ne montre pas ce qu'il prétend montrer – à savoir que tout système dont les variables sont conjointement indépendantes satisfait la condition de Markov causale. Cette thèse est la conclusion de la section à venir, qui vise plus généralement à déterminer comment le résultat 4.2 établi dans Steel (2005) s'articule au résultat classique 4.1 que Steel prétend dépasser.

4.3 Le résultat de Steel et le résultat classique

Pour que Steel montre bien que la condition de Markov causale s'accommode de l'indéterminisme aussi bien du déterminisme (ou, plus précisément, que le résultat classique peut se passer de la restriction au cas déterministe), il est nécessaire que les variables U_i d'un modèle causal fonctionnel soient des variables exogènes du système réel que ce modèle représente. La première sous-section de la présente section vise à déterminer si c'est bien le cas – c'est-à-dire à déterminer quel est le statut de ces variables problémtiques que sont les variables U_i. Plus généralement, la première section est consacrée à analyser le rapport entre le cadre théorique adopté par Steel et le cadre traditionnel que nous avons introduit dans la première section. A l'issue de cette analyse, nous aurons les moyens d'établir la thèse selon laquelle Steel ne montre pas que la clause déterministe du théorème 4.1 est superflue. Ce sera l'objet de la seconde sous-section.

4.3.1 Les variables problématiques d'un modèle fonctionnel causal

Ainsi que nous venons de l'indiquer, la question qui nous occupe dans cette sous-section est de savoir si les variables U_i d'un modèle fonctionnel causal sont des variables exogènes d'un système réel que ce modèle représente. Cette question demande en fait à être précisée. En effet, deux cadres théoriques différents ont été introduits dans ce qui précède :

1. le cadre classique dans lequel on montre que la condition de Markov causale est satisfaite par tout système déterministe S dont les variables exogènes sont conjointement indépendantes. Dans ce cadre, les variables exogènes de S sont les variables exogènes d'un ensemble de variables qui représentent toutes les propriétés observables pertinentes dont l'instanciation dépend de S ;
2. le cadre introduit par Steel, dans lequel il montre que tous les modèles fonctionnels causaux dont les variables U_i sont conjointement indépendantes satisfont la condition de Markov causale.

Nous voulons maintenant décider si Steel est fondé à dire qu'il montre que tout système S sont les variables exogènes sont conjointement indépendantes est déterministe. Il nous faut donc déterminer si les variables U_i d'un modèle fonctionnel causal M_S qui représente toutes les propriétés observables dont l'instanciation dépend de S, sont les variables exogènes de S.

On peut commencer ici par remarquer que Steel manifeste une hésitation au moment de désigner les variables U_i d'un modèle fonctionnel causal. Plus précisément, il introduit les modèles fonctionnels comme des modèles sur « un ensemble de variables endogènes $\mathbf{X} = \{X_1, \ldots, X_n\}$ [...] et un ensemble de *variables exogènes* ou termes d'erreur $\mathbf{U} = \{U_1, \ldots, X_k\}$ »[26] Ce passage trahit une incertitude relative au statut des variables de $\mathbf{U}$ – dont le lecteur aura compris que nous la considérons significative. Mais par ailleurs il suggère un point sur lequel faire porter l'analyse. Ce point est celui de la différence entre variables exogènes et termes d'erreur.

Variables exogènes et termes d'erreur. Pour définir les termes d'erreur et déterminer en quoi ils diffèrent des variables exogènes, il convient de revenir aux ensembles de variables. Plus précisément, deux points doivent être rappelés ici :

- les variables exogènes d'un ensemble de variables $\mathbf{V}$ sont celles des variables de $\mathbf{V}$ qui n'ont pas de cause directe dans $\mathbf{V}$;
- un ensemble de variables $\mathbf{V}$ peut représenter plus ou moins adéquatement un système réel donné, selon que les variables de $\mathbf{V}$ représentent plus ou moins des propriétés observables dont l'instanciation dépend du système.

Cependant, quel que soit le degré d'adéquation de la représentation d'un système S par un ensemble de variables $\mathbf{V}_S = \{V_1, \ldots, V_n\}$, il est toujours possible d'écrire un ensemble de n équations tel que chaque variable V_i figure du côté gauche d'une équation et d'une seule, et y figure comme une fonction f_i de ses parents dans le graphe causal sur $\mathbf{V}_S$ et d'une variable T_i. Il est courant de représenter S au moyen de l'ensemble de ces équations et d'une distribution de probabilités sur l'ensemble $\mathbf{T}$ des variables T_i.[27] Les représentations de ce type sont couramment appelés « modèles fonctionnels causaux»[28], et nous parlerons plutôt de « modèles fonctionnels causaux usuels» afin de les distinguer des modèles fonctionnels causaux de Steel :

Définition 4.6 (Modèles fonctionnels causaux usuels) *Un* modèle fonctionnel causal usuel *sur un ensemble de variables* $\boldsymbol{V} = \{V_1, V_2, \ldots, V_n\}$ *est un couple composé de :*

1. *un ensemble* $\boldsymbol{E}$ *de n équations tel que chaque* V_i *apparaît du côté gauche du signe « =» dans exactement une équation de* $\boldsymbol{E}$*, et y apparaît comme une fonction de ses causes directes dans* $\boldsymbol{V}$ *et d'une variable* T_i *;*

[26] Steel (2005) p. 5. Les italiques sont dans le texte original.
[27] Sur ce point, voir Pearl (2000) pp. 26–27.
[28] Voir en particulier Pearl (2000) p. 27.

2. une distribution de probabilités sur l'ensemble $\boldsymbol{T} = \{T_1, T_2, \ldots, T_n\}$.

Les variables T_i qui figurent dans un modèle causal fonctionnel usuel sont habituellement appelés « termes d'erreur» – et nous nous conformerons à cette habitude.

La façon dont nous venons de présenter les modèles causaux fonctionnels usuels fait apparaître que les variables T_i d'un tel modèle ont une fonction bien précise : permettre une représentation fonctionnelle des relations entre les variables de $\mathbf{V}$. Pour le dire autrement, chaque T_i représente tout ce qui contribue à la détermination de la valeur de V_i mais n'est pas représenté par l'ensemble de causes directes de V_i dans $\mathbf{V}$. Plus précisément et plus concrètement, les variables de $\mathbf{T}$ « représentent, en tas [*in one heap*], les erreurs de mesure, les facteurs omis et quelque élément purement probabiliste qu'il puisse y avoir dans la détermination d'un effet par ses causes»[29].

De cet éclaircissement emprunté à Cartwright, nous tirons deux points non indépendants. En premier lieu, la différence entre variables exogènes et termes d'erreur se comprend comme une différence de nature entre ce que les unes et ce que les autres représentent, et corrélativement comme une différence entre deux visées de la représentation. D'un côté les variables exogènes représentent des familles de propriétés observables et la représentation vise à produire une image adéquate de la réalité représentée : les propriétés représentées sont des propriétés observables identifiées et chaque variable exogène représente exactement une telle famille. De l'autre côté les termes d'erreur représentent des agrégats d'influences de nature diverse et la représentation vise seulement à procurer au modèle proposé la propriété de donner une représentation fonctionnelle des relations entre les variables de $\mathbf{V}$.

En second lieu, il apparaît que le mode de représentation des systèmes réels que constituent les modèles fonctionnels causaux usuels n'a pas le même statut que celui qui sous-tend l'affirmation classique selon laquelle les systèmes déterministes dont les variables exogènes sont conjointement indépendantes satisfont la condition de Markov causale. D'une part, le recours aux termes d'erreur manifeste une prise en compte des conditions concrètes de construction des modèles et d'évaluation de la valeur des variables. Cette prise en compte peut être corrélative du projet de prédire la valeur de certaines variables à partir de la mesure effective de la valeur d'autres variables. D'autre part, le mode de représentation des systèmes réels que nous avons mentionné dans la première section du présent chapitre et relativement auquel l'affirmation classique fait sens reste abstrait.

[29] Cartwright (1999) p. 9.

Variables exogènes ou termes d'erreur ? Maintenant que nous avons établi la différence entre variables exogènes et termes d'erreur, nous pouvons revenir aux variables U_i des modèles fonctionnels causaux de Steel. Le dernier paragraphe fait apparaître que la question de savoir si ces variables sont des variables exogènes ou des termes d'erreur trouvera une réponse en même temps que la question de savoir ce que ces variables représentent. Or, même si Steel ne traite pas cette dernière question pour elle-même, il semble bien que les variables U_i d'un de ses modèles fonctionnels causaux peuvent représenter des influences relevant de deux types bien différents. Pour le comprendre, nous revenons à l'exemple de la voiture quantique. Le modèle fonctionnel causal dont Steel dit qu'il représente le système de démarrage de cette voiture compte deux variables U_i, sur lesquelles nous revenons tour à tour.

De la première de ces deux variables, U_1, nous avons considéré plus haut qu'elle représente des propriétés dont l'instanciation cause la rotation de la clef. Selon cette lecture, il s'agit d'une variable exogène au sens propre que nous venons de définir : une variable qui représente une propriété observable dont l'instanciation dépend du système considéré, mais qui se trouve ne pas avoir de cause parmi les autres variables utilisées pour représenter le système. Cette lecture est corroborée par la lettre du texte de Steel, qui parle pour U_1 – et pour U_1 seulement – de « parent exogène»[30]. Toutefois, il est apparu déjà que la terminologie utilisée par Steel n'est pas toujours usuelle et l'hypothèse selon laquelle U_1 serait un terme d'erreur ne peut pas encore être rejetée définitivement.

Si U_1 est un terme d'erreur, il s'agit du terme d'erreur pour X_1. Dans ce cas, U_1 représente tout ce qui contribue à la détermination de la valeur de X_1 mais n'est pas représenté par les variables de l'ensemble $\mathbf{X} = \{X_1, X_2\}$. Surtout, si U_1 est un terme d'erreur, alors le modèle envisagé par Steel a le même statut que les modèles fonctionnels causaux usuels. En conséquence, il doit compter un terme d'erreur non seulement pour X_1, mais encore pour X_2.

U_2 se présente d'abord comme un tel terme d'erreur. En effet, il est apparu plus haut que U_2 représente le caractère probabiliste de l'action de X_1 sur X_2. U_2 représente donc bien une contribution à la détermination de la valeur de X_2 qui n'est pas représentée par les variables de l'ensemble $\mathbf{X}$. Seulement, U_2 ne représente pas *tout* ce qui contribue à la détermination de la valeur de X_2 mais n'est pas représenté par les variables de $\mathbf{X}$. Plus précisément, elle ne représente qu'un des types d'influences qui sont représentées ensemble par les termes d'erreur au sens usuel de l'expression. Pour cette raison, elle ne joue pas le rôle que jouerait un terme d'erreur pour X_2 dans un modèle fonc-

[30] Steel (2005) p. 13.

tionnel causal usuel sur **X**. En conséquence, U_2 n'est pas un terme d'erreur. Corrélativement, nous n'avons pas de raison de renoncer à l'analyse selon laquelle U_1 est une variable exogène au sens usuel du terme.

Parce que l'exemple de la voiture quantique est choisi par Steel lui-même pour illustrer son analyse et parce que Steel ne fait par ailleurs jamais référence ni à la possibilité d'omettre des variables au moment de construire un modèle ou ni à d'éventuelles erreurs de mesure, nous considérons que les conclusions de l'analyse menée dans le paragraphe précédent peuvent être généralisées. Autrement dit, nous considérons qu'il est toujours le cas que les variables **U** des modèles fonctionnels causaux de Steel ou bien sont des variables exogènes dans l'ensemble de variables choisi pour représenter un système donné, ou bien représentent des influences de l'un des trois types d'influences représentées par les termes d'erreur usuels. Étant donné un modèle fonctionnel causal du type de ceux que Steel considère, l'ensemble de variables **U** qui lui correspond est donc un objet composite, et l'hésitation de Steel au moment où il introduit les variables U_i trouve une explication. Il nous reste maintenant à tirer les conséquences de cette analyse relativement à la portée du résultat établi dans Steel (2005) et, plus précisément, à son rapport avec le théorème 4.1.

4.3.2 Ce que Steel montre (et ce qu'il ne montre pas)

Rappelons pour commencer cette sous-section que le résultat établi par Steel est le suivant : étant donné un modèle fonctionnel causal M défini sur (**X**, **U**), si les variables de **U** sont conjointement indépendantes, alors la réunion de **X** et de l'ensemble des variables de **U** qui sont des causes directes de variables de **X**, satisfait la condition de Markov causale. Une conséquence immédiate de l'analyse menée dans la dernière sous-section est alors la suivante : le résultat établi par Steel n'est pas un résultat de satisfaction de la condition de Markov causale par tout ensemble de variables dont les variables exogènes sont conjointement indépendantes. Mais il convient ici d'être plus précis. Nous dirons alors : le résultat 4.2 a pour contenu la satisfaction de la condition de Markov causale par tout ensemble de variables dont les variables exogènes sont conjointement indépendantes à la condition expresse d'appeler « variables exogènes» d'un système, les variables U_i d'un modèle fonctionnel causal de Steel représentant ce système – et donc de renoncer au vocabulaire et au format de représentation usuels qui ont été examinés dans la précédente sous-section. En conséquence, il est faux que Steel établit que la clause déterministe est superflue dans le théorème 4.1. Corrélativement, il est au moins trompeur de la part de Steel d'annoncer sans qualification qu'il établit que la condition de Markov causale est satisfaite par les systèmes

dont les variables exogènes sont conjointement indépendantes non seulement quand dans le cas déterministe, mais encore dans le cas général.

Nous venons de montrer que l'énoncé du théorème 4.2 ne peut pas être compris comme l'énoncé de la satisfaction de la condition de Markov causale par tout ensemble de variables dont les variables exogènes sont conjointement indépendantes – ni donc, au-delà, par tout système, déterministe ou non, dont les variables exogènes sont conjointement indépendantes. Mais si le théorème 4.2 n'est pas le résultat que Steel annonce avoir montré, il reste possible que Steel (2005) indique les voies d'une preuve de ce résultat. Plus précisément, il se pourrait que la preuve proposée pour le théorème 4.2 se transforme facilement en une preuve de ce que la condition de Markov causale vaut pour tous les ensembles de variables (et, au-delà, pour tous les systèmes) dont les variables exogènes sont conjointement indépendantes. Cette hypothèse est rendue plausible par la modularité de la preuve proposée par Steel pour son résultat et par le fait corrélé que la stratégie que Steel met en oeuvre est similaire à celle que nous avons utilisée pour montrer le résultat classique 4.1 et à celle que Pearl mentionne pour établir un résultat sensiblement différent.[31]

Revenons donc un moment sur la preuve proposée pour le théorème 4.2. Elle consiste essentiellement en ceci : étant donné un ensemble de variables $\mathbf{V}$ tel que la valeur d'un sous-ensemble $\mathbf{W}$ de $\mathbf{V}$ détermine fonctionnellement la valeur de $\mathbf{V}$ tout entier,

1. on montre que le couple composé du graphe G représentant les relations de dépendance fonctionnelle directe entre les variables de $\mathbf{V}$ et de la distribution de probabilités sur $\mathbf{V}$ satisfait la condition de Markov si les variables de $\mathbf{W}$ sont conjointement indépendantes ;
2. on en déduit que tout couple composé de a) le graphe qui est la restriction de G à un ensemble de variables $\mathbf{X}$ contenant toutes les variables de $(\mathbf{V} \setminus \mathbf{W})$ et au moins toutes celles des variables de $\mathbf{W}$ qui sont des parents communs de plusieurs variables dans G et b) la distribution de probabilités sur $\mathbf{X}$, satisfait la condition de Markov.

Il est explicite que l'hypothèse d'indépendance conjointe est mobilisée lors de l'étape 1., d'une façon qui a été décrite précisément dans la première section du présent chapitre. Pour obtenir à l'issue de 2. un résultat qui vaut en cas d'indépendance conjointe des *variables exogènes* (au sens usuel de l'expression) de $\mathbf{V}$, il faut que la valeur de l'ensemble de ces variables exogènes détermine la valeur de $\mathbf{V}$ tout entier. Autrement dit, ce n'est possible que si $\mathbf{V}$, précisément, est déterministe au sens qui a été défini dans la première sous-section du chapitre. Il en découle que la stratégie adoptée par Steel pour

[31] Pearl (2000) p. 30.

montrer le théorème 4.2 ne permet pas de montrer que la condition de Markov causale est satisfaite par tout ensemble de variables, même indéterministe, dont les variables exogènes sont conjointement indépendantes.

Reste qu'il se pourrait que soit vraie la proposition que la condition de Markov causale est satisfaite par tous les systèmes dont les variables exogènes sont conjointement indépendantes, que ces systèmes soient ou non déterministes. Contre cette hypothèse, nous montrons que l'analyse menée dans Steel (2005) ne change rien au fait que l'usine envisagée dans Cartwright (1999) et que nous avons présentée dans le paragraphe 1.3.2.1 ne satisfait pas la condition de Markov causale. Rappelons que cette usine « recourt à un processus véritablement probabiliste »[32] pour produire un composant chimique, que la probabilité que le composant soit produit quand l'usine fonctionne vaut 0,8 et que des produits polluants sont produits exactement quand l'est le composant. On note X, Y et C les variables qui représentent respectivement si le composant est produit, si des produits polluants sont émis et si l'usine fonctionne. Sous l'hypothèse selon laquelle cet ensemble représente toutes les propriétés observables dont l'instanciation dépend du système de production du composant chimique par l'usine considérée, ce système a C pour unique variable exogène. Il en découle que les variables exogènes de ce système sont conjointement indépendantes. Pourtant, nous avons vu que l'ensemble de variables $\{X, Y, C\}$, et donc le système considéré, ne satisfait pas la condition de Markov causale. La raison en est que la valeur de C ne détermine pas fonctionnnellement celle de ses effets – autrement dit : que le système est indéterministe. Si l'on s'en tient à la terminologie usuelle, l'usine imaginée par Cartwright est bien un contre-exemple à la satisfaction de la condition de Markov par tous les systèmes dont les variables exogènes sont conjointement indépendantes. Si Steel réussit à les considérer comme non indépendantes[33], ce ne peut être qu'au prix d'un usage déviant de l'expression « variable exogène».

En définitive, nous avons montré que la proposition selon laquelle la condition de Markov causale est satisfaite par tous les systèmes dont les variables exogènes sont conjointement indépendantes est vraie seulement si on abandonne le sens usuel de « variable exogène ». Plus précisément, il apparaît que l'indépendance conjointe des variables exogènes est une condition suffisante pour la condition de Markov causale seulement si on accepte parmi les variables exogènes des variables qui représentent le caractère probabiliste de l'action de certaines causes sur leurs effets. On peut soutenir qu'il n'y a rien de problématique à les y accepter, et que le remarquer est jus-

[32] Cartwright (1999) p. 7.
[33] Steel (2005) p. 15–16.

tement et exactement l'apport de Steel (2005) au débat sur la condition de Markov causale. Ce point de vue semble d'ailleurs être celui de Steel lui-même, qui écrit : « l'apport fondamental est que les variables exogènes d'un modèle fonctionnel peuvent être interprétées comme représentant soit des causes, soit de l'indéterminisme véritable»[34]. Le problème est que, nous l'avons vu, les modèles fonctionnels causaux usuels comptent déjà des variables qui représentent la façon dont les causes probabilistes agissent sur leurs effets et que ces variables n'ont jamais été appelées « variables exogènes» avant Steel (2005). En conséquence, Steel devrait prendre soin d'une part de distinguer la terminologie et le mode de représentation des systèmes réels qu'il utilise des terminologie et mode de représentation usuels, et d'autre part de ne pas prétendre contribuer directement au débat sur la condition de Markov causale tel qu'il pré-existe à Steel (2005). Il se trouve qu'il ne fait ni l'un, ni l'autre – et c'est précisément ce que nous lui reprochons. Cela n'implique pas que nous considérons que Steel (2005) ne contribue en rien au débat sur la condition de Markov causale, et plus exactement sur son rapport au déterminisme. Positivement, la dernière section du présent chapitre vise à faire apparaître en quoi il y contribue effectivement.

4.4 Contribution de Steel (2005) au débat sur la condition de Markov causale

La critique que nous venons de porter contre Steel (2005) vise essentiellement la façon dont le résultat 4.2 est annoncé et présenté. Ainsi que nous l'avons indiqué déjà, nous ne remettons en cause ni la vérité de ce résultat, ni la correction de la preuve que Steel en donne. Nous ne mettons pas non plus en doute ceci que Steel (2005) contient une proposition originale, qui consiste dans l'utilisation des modèles fonctionnels causaux tels qu'ils sont définis par Steel. La section qui commence vise à évaluer la contribution de cette proposition au débat sur la condition de Markov causale. Nous le faisons en deux temps. Dans une première sous-section, nous examinons pour eux-mêmes les modèles fonctionnels causaux de Steel. Cet examen nous conduit à soutenir qu'ils ne sauraient être acceptés tels qu'ils sont définis par Steel, mais qu'ils suggèrent un raffinement des modèles fonctionnels causaux usuels. La seconde sous-section est consacrée à montrer comment le rapport entre la condition de Markov causale et le déterminisme peut être précisé dans ce cadre théorique nouveau.

[34]Steel (2005) p. 4.

4.4.1 Examen des modèles fonctionnels causaux de Steel

Il est apparu dans la dernière section que les modèles fonctionnels causaux utilisés dans Steel (2005) diffèrent à la fois des modes de représentation abstraits auxquels nous nous sommes référés jusqu'à la fin de la section 4.1 et des modèles fonctionnels causaux usuels que nous avons présentés dans la section 4.3. La façon dont les modèles utilisés par Steel diffèrent des uns et des autres se marque dans les variables U_i auxquels il recourt. Nous avons vu que ces variables représentent pour certaines des familles de propriétés observables, et pour les autres le caractère probabiliste de l'action de certaines causes sur leurs effets. En conséquence, pour certaines elles sont des variables exogènes, et pour les autres elles représentent des influences usuellement représentées par les termes d'erreur.

Cependant nous avons vu que l'ensemble **U** d'un des modèles fonctionnels causaux de Steel n'est pas la réunion de l'ensemble des variables exogènes et de l'ensemble des termes d'erreur usuels pour un ensemble de variables donné. En effet, il est apparu que celles des variables U_i qui représentent des influences usuellement représentées par les termes d'erreur ne représentent pas *toutes* les influences usuellement représentées par les termes d'erreur. Plus précisément, elles ne représentent que les influences de l'un des trois types usuellement pris en charge. Pour cette raison, il nous semble que les modèles fonctionnels causaux de Steel ne constituent pas un format de représentation des systèmes réels qui peut être accepté. De deux choses l'une en effet, pour qui veut être cohérent :

- soit il adopte un mode de représentation abstrait. Dans ce cas, toutes les propriétés observables dont l'instanciation dépend du système considéré sont représentées, *et elles seules le sont*. En particulier, l'aspect probabiliste de l'action de certaines causes sur leurs effets ne l'est pas ;
- soit il adopte un mode de représentation qui prend en compte les conditions concrètes de la construction des modèles et de la mesure des valeurs de variables. Dans ce cas, il convient que soient représentés non seulement l'aspect probabiliste de l'action de certaines causes sur leurs effets, mais encore les causes éventuellement omises des variables représentant des propriétés observables et les éventuelles erreurs de mesure de la valeur de ces variables.

Les modèles fonctionnels causaux de Steel ne relevant ni de l'une, ni de l'autre des deux branches de cette alternative, nous considérons qu'ils ne constituent pas un mode de représentation cohérent, et donc acceptable, des systèmes réels. En conséquence, nous rejetons la proposition de Steel. Ce rejet est le fruit d'une critique dont il nous faut souligner à nouveau qu'elle est distincte

de celle qui a été menée dans la section précédente.

Si le mode de représentation des systèmes réels que Steel définit ne peut être accepté, nous considérons néanmoins qu'il est porteur d'une suggestion intéressante. Pour le comprendre, il convient de revenir sur la différence entre les modèles fonctionnels causaux de Steel et les modèles fonctionnels causaux usuels. Nous avons montré qu'ils diffèrent principalement par ceci que les variables U_i de Steel représentent les influences d'un seul des trois types d'influences usuellement représentées par les termes d'erreur. Ce qui nous a intéressé en cela jusqu'à présent est que, du coup, les variables U_i ne représentent pas *tout* ce que les termes d'erreur usuels représentent. On peut cependant regarder le même fait sous un autre angle. On insistera alors sur ceci que chaque variable U_i ne représente *qu'une* influence, là où un terme d'erreur usuel en représente plusieurs « en tas» – et donc ne représente adéquatement aucune influence réelle. En d'autres termes, si on les considère sous l'angle du mode de représentation, les modèles fonctionnels causaux de Steel se présentent comme plus réalistes que les modèles fonctionnels causaux usuels.

Corrélativement, les modèles fonctionnels causaux de Steel suggèrent de faire évoluer les modèles fonctionnels causaux usuels vers un plus grand réalisme. Plus précisément, Steel (2005) invite à associer à chaque variable de l'ensemble qu'on considère autant de termes d'erreur qu'il existe d'influences réellement distinctes représentées par le terme d'erreur usuel pour cette variable. Les termes d'erreur ainsi conçus peuvent être qualifiés de « réalistes », de même que les modèles fonctionnels causaux qu'ils invitent à définir :

Définition 4.7 (Modèles fonctionnels causaux réalistes) *Un* modèle fonctionnel causal réaliste *sur un ensemble de variables* $\boldsymbol{V} = \{V_1, V_2, \ldots, V_n\}$ *est composé de :*

1. *un ensemble* $\boldsymbol{E}$ *de n équations tel que chaque variable* V_i *apparaît du côté gauche du signe « =» dans exactement une équation de* $\boldsymbol{E}$ *et y apparaît comme une fonction de :*
   - *ses causes directes dans* $\boldsymbol{V}$*,*
   - *un terme d'erreur réaliste* $T_{i,1}$ *représentant les causes de ce que* V_i *représentent qui ne sont pas représentées par les variables de* $\boldsymbol{V}$*,*
   - *un terme d'erreur réaliste* $T_{i,2}$ *représentant les erreurs possibles de mesure de la valeur de* V_i*,*
   - *autant de termes d'erreur réalistes* $T_{i,2+j}$ *qu'il existe de causes directes de* V_i *dans* $\boldsymbol{V}$ *dont l'action sur* V_i *est probabiliste – et qui chacun représente la façon dont la* j*-ième de ces causes agit sur* V_i *;*
2. *une distribution de probabilités sur l'ensemble* $\boldsymbol{T}$ *des termes d'erreur réalistes.*

Ainsi redéfinis, les modèles fonctionnels causaux ont le réalisme des modèles de Steel : chaque terme d'erreur ne représente qu'une influence réelle. Mais, contrairement aux modèles de Steel, ils constituent un mode de représentation cohérent : les termes d'erreur réalistes représentent toutes les influences représentées par les termes d'erreur dans les modèles fonctionnels causaux usuels. Il s'agit plus précisément d'un mode de représentation qui prend en compte les conditions concrètes de la construction du modèle et de l'évaluation des variables – c'est-à-dire la possibilité que des causes soient omises et celle que des mesures soient erronées. Dans ces conditions, contrairement aux représentations dont il a été question dans la première section du chapitre, les modèles fonctionnels causaux réalistes ne sont pas dans une relation de correspondance bi-univoque avec les systèmes qu'ils représentent. Etant donné un système réel donné, un modèle fonctionnel causal réaliste est défini sur un ensemble de variables représentant des propriétés observables dont l'instanciation dépend du système, mais pas nécessairement sur l'ensemble de toutes les variables de ce type.

Dans cette sous-section, nous avons examiné pour elle-même la proposition que constituent les modèles fonctionnels causaux de Steel. Il est apparu que cette proposition ne peut pas être acceptée, mais qu'elle suggère de définir des modèles fonctionnels causaux réalistes, tels que les influences que les termes d'erreur usuels représentent « en tas» soient représentées séparément. Il reste à déterminer ce qu'on gagne à représenter ainsi les systèmes réels. Pourquoi les modèles fonctionnels causaux usuels devraient-ils être rendus réalistes ?

En réponse à cette question, on pourrait avancer que leur plus grand réalisme suffit à justifier qu'on recoure aux modèles fonctionnels causaux réalistes plutôt qu'aux modèles fonctionnels causaux usuels. Toutefois il n'est pas évident que cet argument soit concluant quand ce qu'on rend réaliste est justement la représentation de ce qu'on ne sait pas identifier clairement. Ce qui nous semble concluant, en revanche, est que les modèles fonctionnels causaux réalistes permettent de donner à la question de savoir si le déterminisme est plus favorable que l'indéterminisme pour la satisfaction de la condition de Markov causale, une réponse plus nuancée que la réponse classique selon laquelle l'indépendance conjointe des variables exogènes en est une condition suffisante dans le seul cas déterministe. Nous nous attachons à le montrer dans la dernière sous-section du chapitre.

4.4.2 Modèles fonctionnels causaux réalistes, déterminisme et condition de Markov causale

Modèles fonctionnels causaux réalistes. Avant de montrer comment les modèles fonctionnels causaux réalistes contribuent au débat sur la satisfaction de la condition de Markov causale, nous nous arrêtons un moment sur ces modèles eux-mêmes. Plus spécifiquement, il nous faut commencer par mentionner que ces modèles, de même que les modèles de Steel, entretiennent une « correspondance immédiate»[35] avec les graphes orientés. A un modèle fonctionnel causal réaliste M défini sur un ensemble de variables $\mathbf{V}$ correspond le graphe orienté G_M dont les sommets sont les variables de $\mathbf{V}$ et tous les termes d'erreurs qui apparaissent dans M, et qui est tel qu'il existe une flèche de W vers V_i dans G_M exactement quand W apparaît du côté droit de l'équation de M qui correspond à V_i.[36]

En outre, et encore une fois sous l'hypothèse selon laquelle G_M est acyclique, toute distribution de probabilités p sur l'ensemble $\mathbf{T}$ des termes d'erreur qui figurent dans M s'étend univoquement à une distribution de probabilités sur $(\mathbf{T} \cup \mathbf{U})$.

Enfin on notera, de même qu'un modèle fonctionnel causal de Steel, M détermine univoquement un graphe causal sur $\mathbf{V}$. Nous noterons ce graphe « CG_M ». Il est la restriction de G_M aux variables de $\mathbf{V}$.

Modèles fonctionnels causaux réalistes et condition de Markov causale. Sous les notations introduites dans le dernier paragraphe, on a le résultat suivant :

Théorème 4.3 *Soit M un modèle fonctionnel causal réaliste défini sur $\boldsymbol{V}$ et $\boldsymbol{T}$ l'ensemble des termes d'erreur qui figurent dans M.*
Si les variables de $\boldsymbol{T}$ sont conjointement indépendantes, alors $\boldsymbol{V}$ satisfait la condition de Markov causale.

Ce résultat est l'équivalent, pour le mode de représentation des systèmes réels que constituent les modèles fonctionnels causaux réalistes, du résultat 4.2 de Steel. Il s'établit de façon strictement analogue.

A titre d'illustration de l'utilisation des modèles fonctionnels causaux réalistes et du théorème 4.3, considérons une voiture classique – c'est-à-dire non quantique. Nous supposons que cette voiture démarre à chaque fois qu'on tourne la clef et qu'il y a de l'essence dans le réservoir. On s'intéresse au système constitué par les mécanismes qui régissent le démarrage de cette voiture et on considère l'ensemble de variables $\mathbf{V} = \{V_1, V_2, V_3\}$ qui représentent chacune des propriétés observables dont l'instanciation dépend du système :

[35] Steel (2005) p. 7.

[36] Nous utilisons la lettre W parce que la variable que cette lettre désigne peut être une variable de $\mathbf{V}$ ou un terme d'erreur.

V_1 prend la valeur 1 si la clef est tournée et la valeur 0 sinon, V_2 prend la valeur 1 s'il y a de l'essence dans le réservoir et la valeur 0 sinon, V_3 prend la valeur 1 si la voiture démarre et la valeur 0 sinon. Un modèle fonctionnel causal réaliste M sur $\mathbf{V}$ se compose des équations :

$$\begin{aligned} V_1 &= g_1(T_{1,1}, T_{1,2}) \\ V_2 &= g_2(T_{2,1}, T_{2,2}) \\ V_3 &= g_3(V_1, V_2, T_{3,1}, T_{3,2}) \end{aligned}$$

où chaque $T_{i,1}$ représente les causes omises de V_i, et chaque $T_{i,2}$ représente les erreurs dans la mesure de V_i. Selon le théorème 4.3, une condition suffisante pour que $\mathbf{V}$ satisfasse la condition de Markov causale est que les variables de $\mathbf{T} = \{T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}, T_{3,1}, T_{3,2}\}$ soient conjointement indépendantes.

Si l'indépendance conjointe des variables de $\mathbf{T}$ est suffisante pour que l'ensemble de variables $\mathbf{V}$ considéré dans le paragraphe précédent satisfasse la condition de Markov causale, elle n'est pas nécessaire. On peut en effet l'affaiblir à partir de la remarque suivante, déjà formulée par Cartwright dans un autre contexte : ce que requiert une preuve du type de celle que nous avons détaillée pour le théorème 4.1 n'est pas l'indépendance conjointe des termes d'erreur, mais celle des « effets nets»[37] des termes d'erreur associés à une même variable. Dans l'exemple que nous venons de proposer, ces effets nets sont les effets conjugués des variables des ensembles $\{T_{1,1}, T_{1,2}\}$, $\{T_{2,1}, T_{2,2}\}$, $\{T_{3,1}, T_{3,2}\}$. La condition suffisante pour la condition de Markov causale à laquelle Cartwright fait référence est la suivante : pour tout couple de réunions disjointes d'ensembles de cette liste, les deux ensembles qui composent le couple sont indépendants. Soyons explicites : pour que la condition de Markov causale soit satisfaite par $\mathbf{V}$ il suffit que :

1. $\{T_{1,1}, T_{1,2}\}$, $\{T_{2,1}, T_{2,2}\}$ et $\{T_{3,1}, T_{3,2}\}$ soient deux à deux indépendants ;
2. $\{T_{1,1}, T_{1,2}\}$ et $\{T_{2,1}, T_{2,2}, T_{3,1}, T_{3,2}\}$ soient indépendants ;
3. $\{T_{2,1}, T_{2,2}\}$ et $\{T_{1,1}, T_{1,2}, T_{3,1}, T_{3,2}\}$ soient indépendants et
4. $\{T_{3,1}, T_{3,2}\}$ et $\{T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}\}$ soient indépendants.

Il n'est donc pas nécessaire que les variables de $\{T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}, T_{3,1}, T_{3,3}\}$ soient conjointement indépendantes. En particulier, il n'est pas nécessaire que (pour i donné) $T_{i,1}$ soit indépendant de $T_{i,2}$ pour que la condition de Markov causale soit satisfaite par $\mathbf{V}$.

Ce que nous venons d'affirmer sur un exemple, il faut maintenant de l'établir dans le cas général. En vue de cela, convenons d'une notation : étant donné un modèle fonctionnel causal réaliste M sur l'ensemble de variables

[37] Cartwright (2001) p. 18.

$\mathbf{V} = \{V_1, \ldots, V_n\}$, on notera $\varphi(i)$ le nombre (supérieur ou égal à 2) de termes d'erreur associés à V_i dans M. Sous cette convention, on a le résultat suivant :

Théorème 4.4 *Soit M un modèle fonctionnel causal réaliste défini sur $\boldsymbol{V} = \{V_1, V_2, \ldots, V_n\}$.*
Si pour tout $\boldsymbol{V'} = \{V_i, \ldots, V_j\}$ et tout $\boldsymbol{V''} = \{V_k, \ldots, V_l\}$ sous-ensembles de $\boldsymbol{V}$ non vides et disjoints, $\{T_{i,1}, \ldots, T_{i,\varphi(i)}, \ldots, T_{j,1}, \ldots, T_{j,\varphi(j)}\}$ est indépendant de $\{T_{k,1}, \ldots, T_{k,\varphi(k)}, \ldots, T_{l,1}, \ldots, T_{l,\varphi(l)}\}$,
alors $\boldsymbol{V}$ satisfait la condition de Markov causale.

Dans la fin du texte, l'antécédent de l'implication qui figure dans le théorème 4.4 sera abrégé de la façon suivante : « **V** satisfait la condition **IC**». Par ailleurs, on pourra dire d'un système *tel qu'il est représenté par le modèle fonctionnel causal réaliste M* qu'il satisfait **IC** si l'ensemble de variables sur lequel M est défini satisfait **IC**. La précision que nous donnons en italiques est rendue nécessaire par ceci, que nous avons déjà souligné, qu'il n'existe pas de correspondance bi-univoque entre les systèmes réels et les modèles causaux fonctionnels réalistes qui les représentent. Contrairement à ce qui était le cas pour le mode de représentation abstrait que nous avons considéré dans la première section et sur lequel est fondé le résultat classique 4.1, les modèles fonctionnels causaux réalistes ne permettent pas d'énoncer un résultat de satisfaction de la condition de Markov causale qui porte directement sur les systèmes réels.

Le théorème 4.4 ne peut pas être établi de façon analogue à celle dont s'établissent le théorème 4.3 et le résultat 4.2 de Steel. En effet, sans l'hypothèse d'indépendance conjointe des termes d'erreur, on ne peut pas montrer que le graphe G_M qui correspond à un modèle fonctionnel causal réaliste sur **V** forme avec la distribution de probabilités sur $(\mathbf{V} \cup \mathbf{T})$ un couple qui satisfait la condition de Markov. Positivement, on prouve le théorème 4.4 de la façon suivante :

Preuve : Soit $\mathbf{V} = \{V_1, \ldots, V_n\}$ un ensemble de variables et M un modèle fonctionnel causal réaliste sur **V**. On note GC_M le graphe causal sur **V**, p la distribution de probabilités sur $(\mathbf{V} \cup \mathbf{T})$, et on suppose que **V** satisfait **IC**.
Considérons une variable V_i de **V**. Il faut montrer qu'elle est indépendante, relativement à l'ensemble de ses causes directes dans **V**, de toute variable de **V** qui n'est pas l'un de ses descendants dans GC_M.
La valeur de V_i est fonctionnellement déterminée par celle de ses causes directes dans **V** et des termes d'erreur qui lui sont associés dans M. Ainsi, pour une valeur fixée de l'ensemble de ses causes directes dans **V**, la valeur de V_i ne dépend que de la valeur des termes d'erreur qui lui sont associés dans M. En outre, la valeur de n'importe quelle variable de **V** est déterminée fonctionnellement par celle de

l'ensemble des termes d'erreur associés dans M à ses causes (directes ou non) dans **V**. Pour une variable V_j qui n'est pas un descendant de V_i dans M, les termes d'erreur associés à V_i n'appartiennent pas à l'ensemble des termes d'erreur dont la valeur détermine fonctionnellement celle de V_j. En conséquence, que **V** satisfait **IC** implique que V_i est indépendante de V_j relativement à l'ensemble de ses causes directes dans **V.**

Le théorème 4.4 dépend, pour ce qui est de son énonciation même, des modèles causaux fonctionnels réalistes. Il en découle que si nous montrons que ce théorème permet de préciser le rapport entre le déterminisme et la satisfaction de la condition de Markov causale, nous aurons montré que les modèles fonctionnels causaux réalistes contribuent au débat relatif au rapport entre l'un et l'autre. C'est ce que nous nous attachons à faire dans la fin la présente sous-section.

Condition de Markov causale et déterminisme. En vue de montrer que le théorème 4.4 constitue une avancée sur le terrain de l'explicitation du rapport entre la condition de Markov causale et le déterminisme, il nous faut caractériser d'abord le déterminisme dans le cadre théorique constitué par les modèles fonctionnels causaux réalistes. Plus précisément, et conformément à la thèse que nous avons défendue au début du présent chapitre, il nous faut indiquer comment les modèles fonctionnels causaux réalistes permettent de caractériser les ensembles de variables déterministes. Cette caractérisation est aisée et de même inspiration que la définition que Steel donne des « modèles fonctionnels déterministes»[38] :

Caractérisation 4.8 (Ensemble de variables déterministe) *Un ensemble de variables* ***V*** *est* déterministe *si et seulement si le modèle fonctionnel causal réaliste sur* ***V*** *ne comporte aucun terme d'erreur représentant la façon dont une cause probabiliste agit sur ses effets directs.*

Cette caractérisation n'est rien d'autre que la formulation, dans le cadre théorique constitué par les modèles fonctionnels causaux réalistes, de la définition 4.2 des ensembles de variables déterministes que nous avons donnée plus haut dans ce chapitre. On peut considérer que, de même que la définition 4.2, la caractérisation 4.8 fonde une caractérisation des *systèmes* déterministes. Mais contrairement à celle que fonde la définition 4.2, cette caractérisation des systèmes déterministes est relative à un modèle. La raison en est, redisons-le, qu'il n'existe pas de correspondance bi-univoque entre les systèmes réels et les modèles fonctionnels causaux réalistes. La caractérisation

[38]Steel (2005) p. 9.

qu'on obtient est la suivante : un système réel tel qu'il est représenté par un modèle fonctionnel causal réaliste M est déterministe si l'ensemble de variables sur lequel M est défini est déterministe.

Sous la caractérisation 4.8, le théorème 4.4 implique un sens auquel le déterminisme est plus favorable que l'indéterminisme pour la satisfaction de la condition de Markov causale. Pour le comprendre, reprenons l'exemple de la voiture quantique introduit par Steel. Plus précisément, nous considérons une voiture qui, de la même façon que celle qui est envisagée par Steel, ne démarre qu'avec une probabilité de 0,85 quand toutes les conditions nécessaires au démarrage sont réunies. Autrement dit, nous considérons une voiture qui démarre seulement si la clef est tournée dans le contact et s'il y a de l'essence dans le réservoir, mais avec une probabilité de 0,85 seulement quand ces deux conditions sont réunies.

Le système de démarrage de cette voiture peut être représenté par un modèle fonctionnel causal réaliste M' sur l'ensemble de variables $\mathbf{V} = \{V_1, V_2, V_3\}$ introduit dans le paragraphe précédent. M' se compose des équations suivantes :

$$\begin{aligned} V_1 &= g_1(T_{1,1}, T_{1,2}) \\ V_2 &= g_2(T_{2,1}, T_{2,2}) \\ V_3 &= T_{3,3} \times g_3(V_1, V_2, T_{3,1}, T_{3,2}) \end{aligned}$$

où $T_{3,3}$ représente une variable binaire de valeurs possibles 0 et 1 et telle que $p(T_{3,3} = 1) = 0,85$. Alors il découle du théorème 4.4 que $\mathbf{V}$ satisfait la condition de Markov causale si :

1'. $\{T_{1,1}, T_{1,2}\}$, $\{T_{2,1}, T_{2,2}\}$ et $\{T_{3,1}, T_{3,2}, T_{3,3}\}$ sont deux à deux indépendants ;

2'. $\{T_{1,1}, T_{1,2}\}$ et $\{T_{2,1}, T_{2,2}, T_{3,1}, T_{3,2}, T_{3,3}\}$ sont indépendants ;

3'. $\{T_{2,1}, T_{2,2}\}$ et $\{T_{1,1}, T_{1,2}, T_{3,1}, T_{3,2}, T_{3,3}\}$ sont indépendants et

4'. $\{T_{3,1}, T_{3,2}, T_{3,3}\}$ et $\{T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}\}$ sont indépendants.

Or, si ces conditions sont satisfaites, alors le sont aussi les conditions 1. à 4. mises au jour plus haut dans la présente sous-section. En effet, pour deux ensembles de variables $\mathbf{X}$ et $\mathbf{Y}$, si $\mathbf{X}$ et $(\mathbf{Y} \cup \{T_{3,3}\})$ sont indépendantes, alors sont indépendants aussi $\mathbf{X}$ et $\mathbf{Y}$. Il en découle que M satisfait **IC** si M' satisfait **IC**. La réciproque, en revanche, est fausse.

Ce que nous venons de mettre en évidence ne dépend pas d'une particularité des systèmes de démarrage des automobiles. Au contraire, il apparaît que vaut généralement ceci : étant donnés deux modèles fonctionnels causaux réalistes M et M' définis sur un même ensemble de variables $\mathbf{V}$ et qui ne diffèrent que par ceci que figurent dans M' des termes d'erreur absents de M, si $\mathbf{V}$ satisfait **IC** quand les relations entre les variables de $\mathbf{V}$

sont adéquatement représentées par M', alors **V** satisfait **IC** quand les relations entre les variables de **V** sont adéquatement représentées par M – et la réciproque n'est pas vraie.

Pour le dire autrement, considérons deux systèmes S et S' qui ne diffèrent que par ceci que S est déterministe tandis que S' ne l'est pas. Soit aussi un ensemble de variables **V** qui représente des propriétés observables dont l'instanciation dépend de l'un ou de l'autre de ces systèmes et au nombre desquelles figure (au moins) une propriété qui dont l'instanciation est déterminée dans S mais pas dans S'. Le modèle fonctionnel causal réaliste sur **V** qui représente S est déterministe si le modèle fonctionnel causal réaliste sur **V** qui représente S' est déterministe. En conséquence, la condition suffisante pour la satisfaction de la condition de Markov causale qui a été mise au jour dans cette sous-section est satisfaite par **V** en tant qu'il permet de représenter S, si elle est satisfaite par **V** en tant qu'il permet de représenter S'. Mais, encore une fois, la réciproque n'est pas vraie : il est possible que le modèle fonctionnel causal réaliste sur **V** qui représente S satisfasse la condition **IC** alors que le modèle sur **V** qui représente S' ne la satisfait pas. Dans ce cas, il est possible que la condition de Markov causale soit satisfaite par **V** en tant qu'il permet de représenter S, mais qu'elle ne soit pas satisfaite par **V** en tant qu'il permet de représenter S'. Au sens précis que nous venons de déployer, le déterminisme est plus favorable que l'indéterminisme pour la satisfaction de la condition de Markov causale.

4.5 Conclusion

4.5.1 Conclusion du chapitre

En définitive, le chapitre qui s'achève a fait apparaître que le déterminisme est plus favorable que l'indéterminisme à la satisfaction de la condition de Markov causale – contrairement à ce que soutient Steel dans l'article qui nous a intéressés. Nous avons défini précisément le sens auquel cette proposition est vraie. Nous considérons que définir ce sens et montrer que la proposition est vraie au sens défini contribue au débat sur la condition de Markov causale. Toutefois, il est bien entendu que cette contribution ne clôt nullement le débat sur l'extension du domaine au sein duquel la condition de Markov causale est satisfaite. *A fortiori* ne règle-t-elle pas le problème plus général de l'extension du domaine au sein duquel les inférences causales fondées sur les réseaux bayésiens donnent des résultats corrects.

Pour définir le sens auquel le déterminisme est plus favorable que l'indéterminisme pour la condition de Markov causale, nous avons recouru

aux modèles fonctionnels causaux réalistes introduits au début de la dernière section. Or, ces modèles sont suggérés par la façon dont est représentée dans les modèles causaux fonctionnels de Steel l'action des causes probabilistes sur leurs effets. Dans cette mesure, Steel (2005) contribue au débat sur la condition de Markov causale. Il n'en reste pas moins que, de cet article, nous rejetons ensemble la conclusion selon laquelle le déterminisme n'est pas plus favorable que l'indéterminisme à la satisfaction de la condition de Markov causale, l'hypothèse selon laquelle la vérité du résultat 4.2 reviendrait à la superfluité de la clause déterministe dans le théorème classique 4.1 et la thèse selon laquelle les modèles fonctionnels causaux définis par Steel constitueraient un mode de représentation cohérent des relations causales au sein d'un système réel.

On notera finalement que la dernière des critiques que nous venons de rappeler porte sur les seuls modèles fonctionnels causaux de Steel, dans ce qu'ils ont de particulier. Aussi, nous n'avons pas critiqué la notion générique de modèle fonctionnel causal. Plus, c'est seulement en référence à un certain type de modèles fonctionnels causaux que les définitions de la sous-section 4.4.2 font sens et que le résultat 4.4 se démontre. En acceptant les modèles fonctionnels causaux, nous nous conformons à une pratique aujourd'hui dominante dans la littérature relative à la condition de Markov causale et, du coup, pouvons prétendre contribuer à ce débat. Il n'en reste pas moins qu'on peut légitimement douter de la capacité de ces modèles à représenter des systèmes réels. Ainsi Gillies avoue-t-il « avoir de sérieux doutes quant au fait qu'aucun système de quelque importance puisse être représenté de cette façon »[39] – c'est-à-dire au moyen d'un modèle fonctionnel causal. Surtout, il fait remarquer qu'au long de deux articles conséquents dans lesquels ils utilisent des modèles foctionnels causaux[40], Halpern et Pearl ne donnent qu'un exemple réel – relativement simple – pour quatorze exemples fictifs. Indubitablement, il y a là un point intéressant, mais qui dépasse le cadre du présent travail dans sa forme actuelle.

4.5.2 Conclusion de la partie

Dans cette première partie, nous avons répondu à la question des corrélats épistémologiques des théories probabilistes de la causalité générique. Cette question a été abordée comme celle de la contribution des réseaux bayésiens à l'épistémologie de la causalité générique. Nous avons justifié cette approche dans l'introduction. Le chapitre 2 donne de nouvelles raisons, ou plutôt des

[39]Communication personnelle.

[40]Halpern et Pearl (2005).

raisons mieux établies, pour lesquelles cette approche est bien fondée.

En effet, prenant appui sur l'analyse de la notion de réseau bayésien causal menée dans le chapitre 1, le chapitre 2 a posé la question de la contribution des réseaux bayésiens à l'épistémologie de la causalité générique du point de vue de l'analyse conceptuelle. Plus précisément, il s'est agi de comparer le critère de causalité véhiculée par les réseaux bayésiens causaux aux théories probabilistes de la causalité. S'il est apparu que les deux types d'analyses n'ont pas exactement les mêmes objets, ces objets ne sont pas si distincts qu'on ne puisse réduire leur différence pour comparer effectivement les analyses. Or, il apparaît alors que la caractérisation de la causalité véhiculée par les réseaux bayésiens causaux a sa place – que nous avons située précisément – dans le champ des théories probabilistes de la causalité.

Le chapitre 3 a déterminé si et comment les réseaux bayésiens renouvellent la méthodologie de l'inférence aux causes génériques. A cette question, nous avons apporté une réponse nuancée. Ainsi nous avons montré que les réseaux bayésiens ne renouvellent nos méthodologies d'inférence causale dans le sens radical où l'entendent leurs partisans. Mais nous avons montré qu'ils mettent en lumière des hypothèses traditionnellement tues et, corrélativement, invitent à une redéfinition des procédures d'inférence causale.

Ainsi que nous l'avons expliqué au début du présent chapitre, la question qu'il traite est orthogonale à celles qui sont traitées plus haut dans la partie. En effet, il ne s'agit plus de tirer des conséquences de ce que les hypothèses d'acyclicité et de fidélité et, surtout, la condition de Markov causale ne sont pas satisfaites dans le cas général. Positivement, ce dernier chapitre prend place dans les débats relatifs à l'extension et à la caractérisation du domaine au sein duquel ces hypothèses sont satisfaites. Ainsi que nous l'avons indiqué déjà, ce chapitre est très loin de clore ces débats. Plus, il serait irréaliste de prétendre les clore ici. Dans ces conditions, il est raisonnable que nous nous en tenions là de nos investigations relatives à la causalité générique. Il convient que nous nous tournions vers les questions que les théories probabilistes de la causalité soulèvent relativement à la causalité singulière.

Deuxième partie

Causalité singulière et propensionnisme

Notre travail trouve son unité dans ceci qu'il traite exclusivement de questions que soulèvent les théories probabilistes de la causalité, ou plus exactement qui se posent en l'état actuel de leur développement. C'est ainsi que les théories probabilistes de la causalité générique soulèvent des questions épistémologiques et nous ont conduits à nous intéresser aux réseaux bayésiens en tant qu'ils permettent d'inférer des causes. De leur côté, les théories probabilistes de la causalité singulières existantes laissent ouverte une question d'analyse conceptuelle. Plus précisément, elles laissent ouverte la question de l'analyse du concept de causalité actuelle, c'est-à-dire de la causalité singulière en tant qu'elle se caractérise d'abord par son inscription dans l'espace et dans le temps et qu'elle est une relation entre les contenus de zones d'espace-temps.

La partie qui commence prend place dans le champ théorique ainsi ouvert puisqu'elle traite du rapport entre la causalité actuelle et les probabilités. La question est abordée depuis la philosophie des probabilités, pour la raison stratégique que cela permet de circonscrire à peu de frais un domaine d'investigation. En effet, nous soutenons que c'est seulement sous une interprétation propensionniste que les probabilités sont susceptibles d'entretenir un rapport avec la causalité actuelle. A cette première raison de s'intéresser au propensionnisme vient s'adjoindre une seconde, qui en est indépendante. Il s'agit de ceci que le propensionnisme – au moins sous certaines de ses formulations – est une interprétation sous laquelle les probabilités font référence à des entités causales. Dans le chapitre 5, nous présentons le propensionnisme ; dans le chapitre 6, nous essayons de déterminer quel rapport la causalité et les probabilistés conditionnelles peuvent entretenir dans un cadre propensionniste.

Chapitre 5

La théorie propensionniste des probabilités

Nous venons de rappeler que l'objet du chapitre qui commence est de présenter la théorie propensionniste des probabilités et que cette présentation a deux justifications. Selon la première, le propensionnisme est la seule théorie physique des probabilités singulières. A ce titre, il se présente comme la seule théorie des probabilités qui nous permette de poser la question qui nous occupe dans cette seconde partie : celle du rapport entre les probabilités et la causalité actuelle. Selon la seconde, la théorie propensionniste des probabilités mobilise de manière essentielle des concepts de type causal. Nous insisterons sur cette dimension autant que faire se pourra. Avant d'en venir à la présentation elle-même, il nous faut finalement souligner que la présentation proposée dans le présent chapitre essentiellement sur le propensionnisme historique, celui de Popper.

Au début du dernier paragraphe – et dans le titre du chapitre, nous parlons de « théorie » plutôt que d'« interprétation » propensionniste des probabilités. Par là, nous signalons principalement que nous souhaitons donner à notre présentation une portée large. Pour le dire autrement, la présentation que nous proposons ne porte pas exclusivement sur la question du rapport du propensionnisme au calcul des probabilités. Positivement, le propensionnisme est également envisagé comme une théorie philosophique – ou, à tout le moins, comme une théorie inséparable de certaines options philosophiques.

Ces options philosophiques sont mises au jour et discutées. En particulier, elles sont confrontées aux options philosophiques qu'engagent les autres théories du calcul des probabilités, principalement fréquentiste et subjectiviste. Ces confrontations et, au-delà, la mise au jour des corrélats philosophiques du propensionniste supposent que le propensionnisme a été caractérisé. Dans ces conditions, l'organisation du chapitre qui commence est

simplement la suivante : dans une première section, nous caractérisons le propensionnisme ; dans une seconde section, nous en envisageons les corrélats philosophiques. Dans une courte dernière section, nous ressaisissons les principaux résultats de notre enquête.

5.1 Caractérisation du propensionnisme

Même en se limitant au propensionnisme popperien, caractériser le propensionnisme n'est pas chose aussi aisée qu'on pourrait se l'imaginer d'abord. En effet, il semble ne pas exister *un*, mais *des* propensionnisme*s*. Ainsi l'une des premières remarques de Gillies dans le chapitre de Gillies (2000a) consacré au propensionnisme est la suivante :

> Dans le cas des théories des probabilités considérées jusqu'à présent (théories classique, logique, subjective et fréquentiste), il existait une version plus ou moins canonique ... Ici, nous avons un "ensemble diffus de propositions"[1] qui sont actuellement développées par des philosophes des sciences différents, dans des directions différentes.[2]

Maintenant, l'existence d'une pluralité de propensionnismes découle de la relative indétermination de la caractérisation popperienne du propensionnisme. Plus précisément, il nous semble que la position de Popper n'est ferme que relativement à une caractérisation minimale du propensionnisme, au-delà de laquelle les hésitations et ambiguïtés laissent ouverte la possibilité de plusieurs propensionnismes qui diffèrent dans le détail. Dans ces conditions, une stratégie semble s'imposer pour ce qui est de l'organisation de la présente section : introduire d'abord la caractérisation minimale du propensionnisme héritée de Popper, préciser ensuite quel est dans le détail le propensionnisme qui nous intéresse ici. Dans une troisième et dernière sous-section, nous abordons la question de savoir si le propensionnisme tel que nous l'aurons caractérisé est une interprétation du calcul des probabilités.

5.1.1 Caractérisation minimale du propensionnismme

Pas plus qu'il n'existe, dans le détail, *un* propensionnisme, il n'existe dans l'oeuvre de Popper *une* caractérisation de la théorie propensionniste des probabilités. De Popper (1957) à Popper (1990), les éléments de caractérisation du propensionnisme sont presque aussi divers qu'ils sont nombreux. Cette

[1] L'expression est introduite dans Miller (1994) p. 175. Miller parle plus précisément d'un « ensemble également diffus de propositions se donnant toutes à elles-mêmes le nom d'interprétation propensionniste des probabilités ».

[2] Gillies (2000a) p. 113.

diversité ne relève pas seulement d'une évolution de la pensée de Popper : d'un côté, des passages de textes différents procèdent d'inspirations assez évidemment communes ; de l'autre, il n'est pas rare qu'un même texte envisage plusieurs éléments de caractérisation différents.

Positivement, il nous semble que c'est la notion même de propension qui est plurivoque. Plus précisément, les propensions semblent se caractériser de plusieurs manières hétérogènes et dont les modalités de coexistence ne sont pas toujours claires. Dans ces conditions, la stratégie que nous adoptons dans la section qui commence est d'abord composite. Plus explicitement, nous envisageons tour à tour trois caractérisations des propensions saillantes et autour desquelles toutes les caractéristiques des propensions semblent pouvoir s'ordonner. Chacune des trois premières sous-sections est consacrée à examiner et discuter l'une des ces caractérisations. Des discussions de ces trois caractéristions des propensions nous tirons, dans la quatrième sous-section, une caractérisation minimale des propensions et, avec elle, du propensionnisme.

5.1.1.1 Les propensions comme propriétés d'ensembles de conditions physiques

Cette première caractérisation des propensions ne prend sens qu'à la lumière de l'idée qui est présentée comme fondamentale pour le propensionnisme dans Popper (1959)[3]. Selon cette idée, la probabilité d'un événement dépend (et ne dépend que) des conditions physiques de son engendrement. La probabilité est alors une *propriété* de cet ensemble de conditions. Toutefois, cette propriété ne se manifeste que si l'événement correspondant se réalise ; il s'agit donc d'une propriété dispositionnelle :

> ... cette modification de l'interprétation fréquentiste [celle qui consiste à définir une suite d'événements par ses conditions d'engendrement] conduit presque inévitablement à conjecturer que les probabilités sont des propriétés dispositionnelles desdites conditions [les conditions d'engendrement] – bref, que ce sont des propensions.[4]

Ce passage appelle trois commentaires.

En premier lieu, il fait apparaître que caractériser les propensions comme des propriétés d'ensembles de conditions physiques ne suffit pas à autoriser une distinction entre propensions et probabilités. Sous cette caractérisation seule, en effet, « propensions » ne peut être qu'un nom pour les probabilités

[3]Popper (1959) pp. 34–35 en particulier.
[4]Popper (1983) p. 372.

conçues comme des propriétés dispositionnelles d'ensembles de conditions physiques.

En deuxième lieu, il semble clair que si les propensions considérées comme des propriétés d'ensembles de conditions physiques sont des dispositions, alors ces dispositions se présentent comme des dispositions d'un type particulier. En effet, étant donnée une disposition, on peut généralement définir un ensemble de conditions dont la réalisation assure la manifestation de la disposition. Ainsi la fragilité d'un verre se manifeste-t-elle si le verre est projeté violemment contre un mur. Maintenant, et ainsi que nous l'avons déjà suggéré, la manifestation de la disposition qu'est une propension est l'occurrence de l'événement dont la propension est la probabilité. Or, il n'existe pas d'ensemble de conditions qui suffise à cette manifestation :

> ... pour cette propension, il n'y aura pas de liste de conditions d'arrière-plan telles que si le dispositif [l'ensemble de conditions physiques] y était exposé, cela produirait nécessairement un certain résultat.[5]

Dans son contexte, cette remarque est présentée comme une objection aux théories propensionnistes des probabilités. Nous ne la comprenons pas comme telle ici, où nous n'en sommes qu'à essayer de comprendre ce qu'est la théorie propensionniste des probabilités.

En troisième lieu, le concept d'ensemble de conditions physiques ayant des propensions pour propriétés demanderait à être précisé. Plus précisément, il faudrait préciser de quel type d'ensembles de conditions physiques les propensions sont des propriétés. Pour le comprendre, imaginons un atome d'uranium 239 (donc radioactif) placé dans une boîte. Clairement, dans ce cas, il existe une propension (ou une probabilité, puisque nous avons vu que cela ne fait pas de différence ici) à la désintégration de l'atome. Ce que nous faisons valoir dans le présent paragraphe est qu'il faudrait préciser si cette propension est une propriété du seul atome d'uranium, de l'ensemble de conditions physiques que constitue l'atome dans la boîte, de celui que constitue la pièce dans laquelle se trouve cette boîte en tant qu'elle s'y trouve, de l'univers tout entier... Cette question d'apparence anodine est en fait source d'importants désaccords entre propensionnistes. Aussi ne la traitons-nous pas dans la présente sous-section, qui est consacrée à une première approche du propensionnisme. Nous pouvons néanmoins remarquer que les différents ensembles de conditions physiques dont une propension pourrait être une propriété diffèrent essentiellement par leur extension, et en venir à la deuxième des caractérisations des propensions que nous avons choisi de présenter.

[5] Eagle (2004) p. 379.

5.1.1.2 Les propensions comme entités métaphysiques. Propensions et forces

La thèse selon laquelle les propensions sont des entités métaphysiques est caractéristique des passages dans lesquels Popper rapproche sa notion de propension de la notion newtonienne de force. Les propensions, en effet, « sont aussi "réelles" que des forces ou des champs de forces »[6]. D'une part elles sont inobservables, d'autre part elles sont des principes d'action qui « ont une réalité physique»[7] ; en un mot, elles sont, comme les forces, méta-physiques.

Le rapprochement des propensions et des forces apparaît lui-même le plus souvent quand Popper se collète avec l'objection selon laquelle les propensions seraient des « entités métaphysiques difficilement acceptables »[8] – cela au mauvais sens que « métaphysique » acquiert dans la seconde moitié du dix-huitième siècle : des « entités invisibles, cachées, ou "occultes" »[9]. En effet, la stratégie adoptée par Popper pour parer à l'objection consiste à s'appuyer sur le rapprochement des propensions et des forces pour suggérer que les propensions sont appelées à connaître non seulement une acceptation aussi large, mais sans doute encore une fortune et une fécondité aussi grandes que celles des forces :

> ... l'introduction de la notion de propension équivaut à une nouvelle généralisation de l'idée de force. Et de même que les positivistes rejetaient cette idée au motif qu'elle faisait entrer en physique ce que Berkeley appelait des "qualités occultes", certains à l'heure actuelle refusent les propensions pour les mêmes raisons.[10]

Ce passage rend clair pourquoi Popper affectionne comparer les propensions aux forces newtoniennes. La récurrence du rapprochement contribue pourtant à faire saillir son imprécision. Pour le comprendre, revenons sur l'idée selon laquelle « l'introduction de la notion de propension équivaut à une nouvelle généralisation de l'idée de force ». Deux pages avant son énonciation, il semble que cette idée peut se comprendre en termes d'inclusion de la classe des forces dans celle des propensions. En d'autres termes, les forces seraient des propensions d'un type particulier :

> ... les forces sont des propensions à mettre des corps en mouvement, à accélérer.[11]

Mais il en semble en aller différemment dès la page suivante :

[6]Popper (1990) p. 33.
[7]Popper (1990) p. 33.
[8]Popper (1983) p. 370.
[9]Popper (1990) p. 34.
[10]Popper (1990) p. 35.
[11]Popper (1990) p. 33.

> Lorsque la propension est inférieure à 1, on peut interpréter cela comme dénotant l'existence de forces concurrentes, "tirant" en quelque sorte le phénomène dans des directions opposées, mais sans encore produire ou contrôler un processus effectif.[12]

Les propensions et les forces n'apparaissent plus alors comme des entités du même type : les forces se combinent en propensions ; une propension est une résultante de forces.

La question du rapport exact qu'entretiennent les propensions et les forces se trouve encore obscurcie par la lecture de Popper (1959). Dans ce texte, Popper compare (déjà) les propensions aux forces, mais envisage – ou du moins semble envisager – une distinction nouvelle entre les unes et les autres :

> Le concept de force – ou mieux encore le concept de champ – introduit une *entité* physique dispositionnelle ... pour expliquer les accélérations observables. De même, le concept de propension, ou de champ de propensions, introduit une *propriété* dispositionnelle d'arrangements physiques expérimentaux singuliers – c'est-à-dire d'événements physiques singuliers – pour expliquer les fréquences observables dans des suites de répétitions de ces événements.[13]

Selon cette dernière citation, les propensions et les forces différeraient comme diffèrent des propriétés et des entités.

5.1.1.3 Les propensions comme possibilités

Des trois que nous avons identifiés, cette caractérisation du propensionnisme est sans doute la moins bien représentée dans l'oeuvre de Popper. Elle est peut-être aussi la plus difficile à analyser. Plus précisément, l'idée de décrire les propensions comme des possibilités est traitée de façon complètement différente dans Popper (1990) et dans Popper (1983).

Dans Popper (1990), les propensions sont introduites – à la suite d'une présentation de l'interprétation classique du calcul des probabilités – comme des « possibilités pondérées ... et qui ont une réalité physique »[14]. La combinaison de possible et de réel autour de laquelle cette présentation s'articule est en elle-même problématique. En effet, le propre du possible en tant que possible est de ne pas avoir de réalité actuelle. Popper semble prendre en compte cette difficulté quand il admet que les propensions, dans l'exacte mesure de leur réalité, sont « *plus que de simples possibilités* »[15]. Toutefois, il

[12] Popper (1990) p. 34.
[13] Popper (1959) p. 31. Nous introduisons les italiques.
[14] Popper (1990) p. 33.
[15] Popper (1990) p. 33. Les italiques sont dans le texte original.

n'ébauche même pas une analyse de la tension qui traverse le couple possible-réel, et de son éventuelle résolution dans le cas présent.

En outre, l'idée de « possibilités pondérées » ne va pas de soi non plus. Revenons en effet aux définitions de la possibilité et de la pondération. Dans le cadre de l'interprétation classique du calcul des probabilités à laquelle Popper semble se référer au point du texte qui nous intéresse, une possibilité est un événement qui n'est ni impossible, ni nécessaire. La pondération, quant à elle, consiste à affecter des grandeurs de coefficients en vue de modifier leur influence respective dans le calcul de leur somme. La notion de « possibilités pondérées » semble donc impliquer l'évaluation d'événements possibles et celle de leur importance respective en vue du calcul d'une certaine grandeur par sommation. Quelle est cette grandeur, selon quels critères des événements peuvent être évalués, et dans quelle unité une telle évaluation pourrait être exprimée, c'est ce que le texte de Popper n'élucide pas.

Quoi qu'il en soit des difficultés soulevées par l'expression « possibilités pondérées ... [et qui] ont une réalité physique », il est au moins clair que la notion de possibilité est au coeur de la caractérisation des propensions qui est proposée au début de Popper (1990). Or, il en va tout autrement dans Popper (1983). Dans ce texte, non seulement la possibilité n'est pas considérée comme un élément central de la définition des propensions, mais encore Popper insiste sur l'insuffisance épistémologique de la notion :

> L'estimation de la *mesure* d'une possibilité (c'est-à-dire l'estimation de la probabilité qui lui est associée) présente donc toujours un aspect prédictif, alors qu'il ne serait guère raisonnable de prédire un événement en sachant seulement qu'il est possible.[16]

La notion qui permet de faire des prédictions, celle qui est scientifique pour Popper est la notion de « *mesure* d'une possibilité ». Or, une « *mesure* de possibilité » ne semble pas être autre chose qu'une propension :

> Il n'est donc pas possible de contourner le fait que nous traitons les mesures de possibilités comme des dispositions, tendances ou propensions.[17]

Dans ces conditions, les propensions sont des grandeurs qui mesurent des possibilités. En ce sens, on peut les considérer comme des probabilités. Il apparaît que cette caractérisation ne permet, pas plus que la première, de distinguer clairement entre les concepts de propension et de probabilité.

La discussion qui s'achève ici n'autorise pas de conclusions fermes. Elle nous a permis simplement de donner au lecteur une idée de la plurivocité du

[16]Popper (1983) p. 371.
[17]Popper (1983) p. 371.

terme « propension », ainsi que des incertitudes conceptuelles et imprécisions terminologiques qui l'entourent. Aussi important soit-il de rendre compte de la richesse de la notion, il nous maintenant à dépasser l'impression de plurivocité, voire de confusion. En effet, il nous faut proposer *une* caractérisation des propensions. Nous considérons que ce que nous a donné à voir le survol qui s'achève ne l'interdit pas. Pour le dire autrement, nous soutenons que l'ensemble des affirmations popperiennnes relatives aux propensions n'est pas incohérent, et qu'on peut caractériser les propensions de façon à faire droit aux intuitions sous-jacentes à chacune des trois caractérisations que nous avons identifiées. Nous nous y essayons maintenant.

5.1.1.4 Caractérisation minimale des propensions

Nous proposons de considérer qu'à chaque ensemble de conditions physiques correspond un ensemble de propensions. L'ensemble de propensions est une propriété de l'ensemble de conditions physiques au sens où il est déterminé par lui. Par ailleurs, les propensions sont des entités métaphysiques précisément au sens où elles sont inobservables en même temps que douées d'une forme de réalité physique. Cette réalité consiste, pour chacune, dans son action en vue de la réalisation d'un événement possible et, au-delà, dans sa capacité à engendrer un (des) phénomène(s) observable(s). La classe des événements possibles est elle-même caractéristique d'une situation physique donnée ; chacun de ces événements peut donc être pensé comme une propriété de cette situation. Cette propriété est dispositionnelle au sens où seule la réalisation d'un événement manifeste qu'il était possible. Cette réalisation manifeste aussi l'existence d'une propension qui y tendait ; en ce sens les propensions sont elles aussi des dispositions. Le degré de réalité des propensions est variable, et mesuré, pour chacune, par la probabilité de l'événement qu'elle tend à actualiser. Par le truchement des propensions, les probabilités sont elles aussi des propriétés d'ensembles de conditions physiques.

Il nous semble que cette caractérisation des propensions et, par voie de conséquence, du propensionnisme ménage une place à chacun des éléments de caractérisation dont nous avons fait état. Plus précisément, la caractérisation que nous proposons est la plus simple parmi celles qui ordonnent de manière intelligible les différentes dimensions que nous avons mises en évidence et discutées. Pour ce qui est de l'intelligibilité, deux points essentiels sont la capacité de la caractérisation retenue à distinguer entre propensions et probabilités d'une part, et entre propensions, probabilités et événements possibles d'autre part. Enfin, la description proposée peut se réclamer de Popper :

> Je me propose d'interpréter la probabilité objective d'un événement isolé comme la mesure d'une *propension* objective – de la force

> d'une tendance, inhérente à la situation physique donnée, à produire l'événement en question, à le faire arriver.[18]

Parce qu'elle ne trahit pas la pensée de Popper et, en même temps, s'est présentée comme l'ordonnancement le plus simple des thèmes saillants dans les multiples caractérisations popperiennes des propensions, nous la considérons comme le socle commun à partir duquel la diversité des propensionnismes s'est développée.

Avant de nous pencher sur cette diversité, il convient d'insister sur les aspects par lesquels le propensionnisme ainsi caractérisé minimalement est une théorie des probabilités qui mobilise de manière cruciale des concepts causaux. Ces aspects, au nombre de deux, sont en fait deux raisons non indépendantes de considérer que les propensions sont de nature causale. La première de ces raisons tient à ce que les dispositions sont au moins de bonnes candidates au titre de causes de leurs manifestations.[19] Il en découle que les propensions, en tant que propriétés dispositionnelles, sont au moins de bonnes candidates au titre de causes. La seconde raison que nous avons de considérer que les propensions sont de nature causale est la suivante : sous la caractérisation que nous venons d'en proposer, les propensions sont des entités méta-physiques capables de produire des événements. Sous cette description, elles se présentent comme des causes au sens où un réaliste peut les envisager. De façon plus générale, il apparaît que la dimension causale appartient au noyau dur du propensionnisme popperien, au sens où elle est impliquée par ce qui le caractérise minimalement.

Parmi les questions que la caractérisation minimale que nous venons de donner laisse ouvertes, la plus remarquable et la plus importante du point de vue conceptuel est celle de la nature de ce que les propensions tendent à réaliser. Plus précisément, il n'est pas clair si les propensions tendent à réaliser des fréquences relatives, ou si elles tendent à réaliser des événements singuliers. De cette relative indétermination de la position popperienne procède la distinction fondamentale de deux propensionnismes : le propensionnisme de long terme d'une part, le propensionnisme de cas singuliers de cas singuliers de l'autre. Décrire le détail du propensionnisme qui nous intéresse dans cette partie de notre travail revient très largement à déterminer s'il s'agit d'un propensionnisme de long terme ou de cas singuliers. Nous nous y attachons dans la prochaine sous-section. Plus spécifiquement, cette sous-section est consacrée à défendre le propensionnisme de cas singuliers, contre le propensionnisme de long terme.

[18] Popper (1983) p. 406.

[19] Pour une défense de la thèse selon laquelle les dispositions sont des causes, voir Mumford (1998) chap.6.

5.1.2 Défense du propensionnisme de cas singuliers

Ainsi que nous venons de l'indiquer, les propensionnistes de long terme (au premier chef : Gillies et Hacking) et les propensionnistes de cas singuliers (Fetzer, Miller, Giere ...) s'opposent sur la question de savoir si les propensions tendent à réaliser des fréquences relatives – ou peut-être plus précisément des « fréquences qui sont approximativement égales aux probabilités »[20] – ou des événements singuliers. Le lecteur aura compris que nous nous rangeons à une théorie de cas singuliers. C'est en effet à la seule condition de considérer que le propensionnisme est une théorie des probabilités singulières qu'il pourra jouer effectivement le rôle que nous prétendons lui faire jouer dans cette seconde partie de notre travail.

Cela, toutefois, ne saurait suffire à justifier que le propensionnisme est bien une théorie des probabilités singulières. Aussi il nous faut donner de bonnes raisons de penser que le propensionnisme doit être considéré comme une théorie des probabilités singulières, ainsi que nous l'avons laissé entendre, plutôt que comme une théorie de long terme. C'est ce à quoi nous nous attelons dans la présente sous-section. Pour commencer, explicitons la distinction entre propensionnisme de long terme et propensionnisme de cas singuliers.

5.1.2.1 Propensionnismes de long terme et de cas singuliers. Deux distinctions secondaires

Ainsi que nous l'avons déjà indiqué, les propensionnismes de long terme et de cas singuliers s'opposent d'abord sur le point de savoir quels types d'événements les propensions tendent à réaliser. Or, à cette distinction fondamentale, Gillies (2000a) en subordonne deux. Ce sont ces deux distinctions secondaires que nous présentons maintenant.

Ensembles de conditions physiques répétables et non répétables. La première de ces distinctions porte sur un point dont précisément nous avons souligné plus haut (à la fin du paragraphe 5.1.1.1) qu'il restait mal déterminé dans le propensionnisme minimal. Ce point est celui de la nature des ensembles de conditions physiques dont les propensions sont des propriétés. Pour les propensionnistes de long terme, il doit s'agir d'ensembles de conditions physiques *répétables*. C'est là en effet une condition nécessaire pour que ces ensembles de conditions physiques engendrent des suites d'événements, et avec elles des « fréquences qui sont approximativement égales aux probabilités ».

[20] Gillies (2000a) p. 126.

Du côté, maintenant, des propensionnismes de cas singuliers, on ne fait pas peser l'exigence de répétabilité sur les ensembles de conditions physiques qui déterminent les propensions. La question qui se pose, et sur laquelle nous aurons à revenir, est plutôt celle de l'extension à donner à ces ensembles. En accord avec la position défendue dans Popper (1990), la plupart des propensionnistes de cas singuliers considèrent qu'il convient de lui donner une extension spatiale maximale, d'en faire « la *situation physique globale* »[21] à un instant donné – ce qui exclut clairement la répétabilité.

Énoncés probabilistes testables et non testables. La seconde des distinctions que Gillies subordonne à la distinction entre propensions à produire des fréquences relatives et propensions à produire des événements singuliers a trait au statut des assignations de probabilités :

> Si les propensions sont assignées à un ensemble de conditions répétables, alors on peut, en répétant les conditions, obtenir des fréquences utilisables pour tester les assignations de propensions. Si, d'un autre côté, les propensions sont attribuées à la "situation complète de l'univers ... au moment considéré", il est difficile, au vu du caractère unique et non répétable de la situation, de voir comment de telles assignations de propensions pourraient être testées.[22]

Selon cette analyse, une théorie de cas singuliers a pour corrélat l'impossibilité de tester les énoncés probabilistes. De l'autre côté, si les propensions sont attachées à des ensembles de conditions répétables et sont des propensions à produire des fréquences relatives, alors les énoncés probabilistes peuvent être soumis à des tests fréquentiels. Sous une conception popperienne de la démarcation entre science et métaphysique, passer d'une théorie de long terme à une théorie de cas singuliers reviendrait donc à « transformer la théorie propensionniste d'une théorie scientifique à une théorie métaphysique »[23].

A ce point, il apparaît que le prix à payer pour endosser une théorie probabiliste de cas singuliers est l'impossibilité de tester les énoncés probabilistes. Pour deux raisons distinctes, nous sommes prêts à accepter ce prix. La première de ces raisons nous est suggérée par Gillies :

> Maintenant, il n'y a rien de mal à développer une théorie métaphysique des propensions, et une telle théorie pourrait être pertinente pour la

[21] Popper (1990) p. 39.
[22] Gillies (2000a) p. 127.
[23] Gillies (2000a) p. 127.

> discussion de vieilles questions métaphysiques, comme le problème du déterminisme.[24]

Ainsi nous semble-t-il que l'impossibilité de tester les énoncés probabilistes n'est pas rhédibitoire dans un contexte où nous poursuivons le but d'analyser le rapport entre les concepts de probabilité et de causalité. La seconde des raisons pour lesquelles nous acceptons de payer le prix de la non-testabilité des énoncés probabilistes pour disposer d'une théorie propensionniste de cas singuliers est que ce prix n'est pas définitif. En d'autres termes, nous soutenons que le propensionnisme de cas singuliers n'est pas exclusif de toute forme de testabilité des énoncés probabilistes. Nous le montrons maintenant, à partir d'une analyse de la position de Popper.

5.1.2.2 Propensionnisme de cas singuliers et testabilité des énoncés probabilistes

Si la position de Popper s'avère particulièrement intéressante pour nous ici, c'est que Popper semble tenir ensemble les deux bouts de la chaîne – le propensionnisme de cas singuliers d'un côté :

> il existe, en général, *inhérente à chaque possibilité* et à chaque lancer, une tendance ou propension à réaliser un certain événement[25],

et la testabilité des énoncés probabilistes de l'autre côté :

> Existe-t-il une valeur qui nous permettrait d'attribuer des valeurs numériques à des possibilités inégales ?
> La réponse qui s'impose est : *oui*, une méthode statistique ; *oui*, pourvu que nous puissions reproduire à loisir la situation qui engendre les événements probabilistes en question, comme dans le cas du dé ; ou que du moins que les événements eux-mêmes se répètent sans que nous n'intervenions, comme dans le cas de la pluie et du beau temps.[26]

Avant d'essayer de comprendre comment Popper articule un propensionnisme de cas singuliers avec la thèse de la testabilité des énoncés probabilistes, insistons sur ceci que c'est bien de cette articulation que témoigne la conjonction des deux passages que nous venons de citer. Pour le dire autrement, la tension entre ces deux passages ne se résout :

- ni en une évolution de la pensée de Popper, chez qui le propensionnisme de long terme des années 1950 aurait progressivement cédé la place à un propensionnisme métaphysique. Le lecteur aura en effet remarqué que les passages cités sont tous deux tirés de Popper (1990) ;

[24] Gillies (2000b) p. 824.
[25] Popper (1990) p. 32. Les italiques sont dans le texte original.
[26] Popper (1990) p. 31.

– ni en une ambiguïté, sinon peut-être une incohérence, de la position de Popper au moment où il écrit les lignes que nous citions. En effet, autant on peut considérer avec Gillies que :

> ... la théorie propensionniste originelle de Popper était, en un sens, à la fois de long terme et des cas singuliers[27],

autant il semble indiscutable que le propensionnisme de Popper (1990) est un propensionnisme de cas singuliers.

Ainsi, les deux passages cités dans le dernier paragraphe indiquent bien que, pour Popper, le propensionnisme de cas singuliers n'est pas exclusif de la testabilité des énoncés probabilistes. Il nous reste à comprendre comment les deux positions peuvent être articulées.

D'après ce que nous avons montré dans le paragraphe 5.1.2.1, la réponse à la question que nous abordons maintenant doit se trouver du côté de la définition des ensembles de conditions physiques dont les propensions sont des propriétés. Nous avons déjà indiqué que Popper écrit dans (1990) que les propensions sont relatives à « la *situation physique globale* »[28] à un instant donné. Dans ce même texte, pourtant, Popper envisage de tester des énoncés probabilistes relatifs à la météorologie de Tel-Aviv en mesurant les fréquences relatives des jours de pluie et des jours de beau temps dans cette ville. Il nous semble que cela n'est possible qu'à la condition suivante : considérer que, relativement à certains événements qu'elle est susceptible d'engendrer, la situation physique globale à un instant donné est suffisamment caractérisée par un ensemble répétable de conditions satisfaites dans la situation. Ainsi, relativement aux phénomènes météorologiques sur Tel-Aviv, la situation physique globale à un instant donné est suffisamment caractérisée par l'ensemble des conditions qui se répètent de jour en jour. Il en découle que les énoncés portant sur des probabilités d'événements météorologiques singuliers peuvent être testées au moyen des fréquences relatives dans la suite des observations météorologiques quotidiennes. Plus généralement, il apparaît que certains énoncés portant sur les propensions de cas singuliers peuvent être testés par les fréquences relatives dans des suites de résultats engendrées par des ensembles de conditions répétables. En pratique, le propensionnisme de cas singuliers n'est donc pas toujours exclusif de la possibilité de tester des énoncés probabilistes au moyen de fréquences observées.

Un point reste cependant à élucider : comment les fréquences relatives dans des suites d'événements engendrés par le même ensemble de conditions physiques pourraient-elles servir à tester des énoncés probabilistes singuliers ? La question est alors celle du lien qui unit les probabilités aux fréquences re-

[27] Gillies (2000a) p. 126.
[28] Popper (1990) p. 39.

latives quand les propensions ne tendent pas à engendrer précisément des fréquences relatives. Ce lien, nous semble-t-il, n'a qu'un seul nom : loi des grands nombres. Selon la loi des grands nombres, les fréquences relatives dans une suite d'événements indépendants résultant de la répétition d'une même expérience aléatoire tendent en probabilité vers les probabilités quand la longueur de la suite tend vers $+\infty$. Dans le cas d'un événement relativement auquel la situation physique globale à un instant donné est suffisamment caractérisée par un ensemble de conditions répétables, les fréquences relatives dans une suite d'événements engendrés par cet ensemble tendent en probabilité vers les probabilités.

Nous venons d'expliquer que la loi des grands nombres établit un lien entre d'une part les probabilités considérées comme des mesures de propensions à réaliser des événements singuliers et d'autre part les fréquences relatives. Ce lien, toutefois, n'est pas tel que les fréquences observées dans des suites d'événements engendrés par un même ensemble de conditions physiques permettraient de tester les hypothèses probabilistes. En premier lieu, en effet, la loi des grands nombres n'énonce pas l'égalité des fréquences relatives et des probabilités, mais seulement la *convergence* des premières vers les secondes. La mise à l'épreuve d'hypothèses probabilistes suppose donc la réalisation de suites infinies d'expériences. En second lieu, la convergence des fréquences relatives vers les probabilités est une convergence *en probabilité*. Il en découle qu'une hypothèse probabiliste qui serait vraie a toujours une probabilité non nulle d'être invalidée par les fréquences statistiques dans une suite finie d'événements engendrés par un même ensemble de conditions physiques. De façon générale, même dans le cas d'un ensemble de conditions répétables, les fréquences observées ne permettent pas de tester les hypothèses probabilistes. La difficulté est exactement la même que celle que nous discutions dans le paragraphe 3.3.2.

Popper connaît bien les deux difficultés que nous venons de présenter, qu'il discute dès *La logique de la découverte scientifique*[29]. Il considère qu'elles peuvent être dépassées en pratique. De longues suites finies d'événements sont alors considérées comme de bonnes approximations de suites infinies. Surtout, des critères de « falsifiabilité méthodologique »[30] sont introduits, qui prennent en compte ceci que

> la distribution de fréquence d'une telle série [série de répétitions indépendantes d'une expérience aléatoire] est en principe "normale"

[29] Popper (1934) chap. VIII.

[30] Gillies (2000a) est particulièrement éclairant sur ce point ; voir Gillies (2000a) pp. 145 et suivantes.

ou "gaussienne".[31]

Plus précisément, ces critères sont l'équivalent, pour le cas particulier d'une distribution normale, des critères de réfutation méthodologique utilisés en statistiques et que nous avons mentionnés dans le paragraphe 3.3.2. Ils permettent de considérer que certaines fréquences relatives observées dans de longues suites finies d'événements engendrés par un même ensemble de conditions physiques réfutent certaines hypothèses probabilistes. Les fréquences relatives permettent donc de tester certaines hypothèses relatives aux mesures de propensions.

En conclusion, nous avons montré dans ce paragraphe comment, en pratique, le propensionnisme de cas singuliers est compatible avec la testabilité de certains énoncés probabilistes. En pratique, et relativement à certains événements singuliers qu'elle a une propension à engendrer, la situation physique globale peut être suffisamment caractérisée par un ensemble de conditions physiques répétables qu'elle satisfait. En pratique, certaines observations de fréquences relatives permettent de réfuter certaines hypothèses relatives aux probabilités singulières. L'analyse des positions de Popper a donc fait apparaître que l'opposition du propensionnisme de long terme au propensionnisme de cas singuliers ne demande pas à être aussi tranchée en pratique que nous avons envisagé qu'elle le soit en théorie.

A ce point, nous avons montré que le problème de la testabilité des énoncés probabilistes n'est pas de nature à invalider le propensionnisme de cas singuliers. D'abord, la raison pour laquelle nous nous intéressons au propensionnisme de cas singuliers n'est pas telle qu'il serait indispensable que les énoncés probabilistes soient testables. Ensuite, l'analyse théorique selon laquelle le propensionnisme de cas singuliers rend non testables les énoncés probabilistes demande à être nuancée : en pratique, certains énoncés relatifs aux probabilités d'événements singuliers peuvent être testés au moyen des fréquences relatives dans de longues suites d'événements engendrés par un ensemble de conditions physiques répétables. Cela, toutefois, ne suffit pas à garantir que le propensionnisme de cas singulier est tenable. En effet, certains auteurs ont soutenu que la notion même de probabilité objective d'un événement singulier est dépourvue sinon peut-être de sens, à tout le moins d'intérêt. C'est à leurs attaques qu'il nous faut répondre maintenant.

5.1.2.3 Probabilités objectives d'événements singuliers

Dans le paragraphe qui commence, nous envisageons les objections aux probabilités objectives d'événements singuliers qui ne portent pas sur l'in-

[31] Popper (1983) p. 303.

terprétation des probabilités conditionnelles. Le problème que les probabilités conditionnelles pose aux théories propensionnistes sera traité plus loin, et extensivement. Il nous reste alors deux critiques principales : d'une part celle qui est développée dans Gillies (2000a) et Gillies (2000b), d'autre part celle qui est présentée dans Kyburg (2002). Nous les présentons chacune à leur tour, puis leur adressons une réponse commune.

La critique de Gillies. La critique de Gillies à l'encontre des probabilités objectives d'événements singuliers repose sur ceci : la probabilité qu'on attribue à un événement singulier dépend d'un grand nombre de facteurs subjectifs. En particulier, elle dépend des classes d'événements pour lesquelles on dispose de données statistiques et de celle de ces classes à laquelle on assigne l'événement singulier qu'on considère.

Du problème posé par l'assignation d'un événement à une classe de référence, Gillies conclut en premier lieu que :

> on devrait considérer que les probabilités dépendent des propriétés utilisées pour décrire un événement, plutôt que de l'événement lui-même.[32]

Autrement dit, il n'existe que des probabilités de classes de référence – c'est-à-dire d'ensembles de propriétés – et ce n'est qu'au prix d'un abus de langage que nous parlons de probabilités d'événements singuliers.

En outre, et c'est là la seconde conclusion de Gillies, une telle probabilité d'événement singulier n'est pas une probabilité objective :

> La procédure générale d'assignation de probabilités à des événements singuliers [...] engage de nombreux éléments subjectifs et, en conséquence, n'est pas susceptible, dans la plupart des cas, de produire une probabilité singulière objective.[33]

Selon Gillies, les probabilités objectives sont les fréquences relatives dans des suites de répétitions d'une expérience aléatoire caractérisée par un ensemble de conditions physiques répétable. Même évaluée en référence à une suite de ce type, une probabilité d'événement singulier n'échappe jamais au subjectivisme.

La critique de Kyburg. La critique développée dans Kyburg (2002) se présente comme visant spécifiquement les théories propensionnistes des probabilités singulières. Toutefois, il nous semble que, à l'instar de celle de Gillies, elle porte plus généralement sur l'idée même de probabilité objective

[32] Gillies (2000b) p. 813.
[33] Gillies (2000b) p. 817.

d'un événement singulier. En effet, Kyburg défend la thèse selon laquelle les fréquences relatives sont les seules probabilités objectives dont nous ayons besoin.

Plus précisément, Kyburg montre que les probabilités objectives sur lesquelles nos utilisations des probabilités sont fondées peuvent toujours être interprétées comme des fréquences relatives – plutôt que comme des probabilités singulières. La stratégie argumentative adoptée par Kyburg est énumérative : 1) il envisage plusieurs situations dans lesquelles les probabilités portent sur des événements clairement singuliers – contrats d'assurance, erreurs de mesure, lancer d'une pièce neuve et destinée à une destruction dès la fin de l'expérience – et 2) pour chacune de ces situations, il montre que nos probabilités évidentielles ne demandent pas à être fondées sur des probabilités objectives autres que les fréquences relatives.

Réponse aux critiques de Gillies et de Kyburg. Aussi différentes soient-elles, les deux critiques dont nous venons faire état sont susceptibles de réponses de même inspiration. Pour le comprendre, nous revenons d'abord à la critique de Gillies. Ainsi qu'il sera apparu lors de la présentation de l'argument de Gillies, sa critique repose tout entière sur l'adoption d'un point de vue épistémique sur les probabilités d'événements singuliers. C'est bien en effet de ce point de vue que se pose le problème de la classe de référence et qu'apparaissent les difficultés qui lui sont apparentées. Dans ce cadre épistémique, Gillies fait apparaître que l'évaluation des probabilités d'événements singuliers dépend d'une multiplicité d'éléments subjectifs :

> A titre d'illustration, supposez que la seule information pertinente de M. B sur M. A est que M. A est un Anglais de 40 ans. Supposez que M. B a une bonne estimation (disons, p) de la probabilité objective que les Anglais de 40 ans vivent jusqu'à 41 ans. Dans ce cas, il serait raisonnable que M. B établisse à p son coefficient de pari subjectif sur la survie de M. A jusqu'à 41 ans, fondant par là objectivement sa probabilité subjective. Cela toutefois ne fait pas de la probabilité subjective de M. B une probabilité objective, car considérez M. C, qui sait que M. A fume deux paquets de cigarettes par jour et qui par ailleurs a une bonne estimation (disons, q) de la probabilité que les Anglais de 40 ans qui fument deux paquets de cigarettes par jour vivent jusqu'à 41 ans. M. C établira sa probabilité subjective concernant le même événement (la survie de M. A jusqu'à 41 ans) à une valeur q différente de la valeur p de M. B.[34]

[34] Gillies (2000b) p. 814.

Ce passage fait apparaître ce que « subjectif » veut dire dans le cadre épistémique adopté par Gillies : est subjectif tout élément qui varie avec les individus – et en particulier avec les connaissances de ces individus. A l'inverse, l'objectivité est la qualité de ce à propos de quoi les jugements individuels concordent.

La notion d'objectivité mobilisée par Kyburg est différente de celle que nous avons mise au jour chez Gillies. Pour Kyburg, en effet, les probabilités objectives sont toujours des « probabilités empiriques objectives »[35] et, corrélativement, sont objectives au sens où elles dépendent de « l'état du monde »[36]. Pour distinguer ces probabilités objectives de celles que Gillies envisage, nous pourrions parler dans ce paragraphe de probabilités *physiques*. Maintenant, toutes physiques soient-elles, ces probabilités sont envisagées seulement en tant qu'elles sont objets de connaissance et peuvent, du coup, contribuer à déterminer « la valeur de la probabilité d'une affirmation A par un agent »[37]. La question des probabilités physiques d'événements singuliers est encore une fois subordonnée à des considérations de nature épistémique.

En définitive, il apparaît que ni Gillies, ni Kyburg ne pose pour elle-même la question qui intéresse le propensionnisme de cas singuliers – c'est-à-dire la question des probabilités physiques d'événements singuliers. En particulier, ni Gillies, ni Kyburg n'établissent que ces probabilités n'existent pas, ou qu'il n'y a pas de sens à parler de probabilité physique d'un événement singulier. Dans ces conditions, nous ne voyons pas que leur critique nous donne de bonnes raisons de considérer que le propensionnisme de cas singuliers n'est pas tenable. De façon exactement similaire, nous avons montré plus haut que l'objection selon laquelle le propensionnisme de cas singuliers serait incompatible avec la testabilité des énoncés probabilistes n'est pas dirimante.

Ainsi avons-nous envisagé et rejeté les deux principales objections au propensionnisme de cas singuliers. Il nous reste maintenant à donner en sa faveur des arguments positifs – et, toujours, indépendants de ceci que le propensionnisme doit être de cas singuliers pour intéresser le projet qui est le nôtre dans cette seconde partie de notre travail. Plus précisément, le dernier paragraphe de la présente section est consacré à présenter les raisons qu'il y a de préférer le propensionnisme de cas singuliers au propensionnisme de long terme.

[35]Kyburg (2002) p. 10.
[36]Kyburg (2002) p. 10.
[37]Kyburg (2002) p. 10.

5.1.2.4 Pour le propensionnisme de cas singuliers, contre le propensionnisme de long terme

Le cas singulier comme raison d'être de la théorie propensionniste des probabilités. Selon le premier de nos arguments, c'est en tant qu'il est de cas singuliers que le propensionnisme a sa place dans le champ des théories des probabilités. Plus précisément, la théorie propensionniste est introduite pour penser des probabilités singulières qui soient objectives (au sens usuel que ce dernier adjectif a dans Kyburg (2002)). D'un côté, en effet, les théories qui rendent compte de manière satisfaisante des probabilités singulières sont des théories épistémiques : les probabilités sont considérées comme des degrés de croyance, et non comme des propriétés du monde physique. De l'autre côté, la seule théorie objective disponible jusqu'au milieu des années 1950 – le fréquentisme – ne permet pas de penser les probabilités d'événements singuliers. Pour le comprendre, il convient de rappeler que, sous une interprétation fréquentiste, une probabilité est toujours celle d'un attribut dans une suite[38] d'événements. Il en découle que les probabilités sont attribuées à des événements génériques, c'est-à-dire à des conjonctions de propriétés.

Maintenant, on pourait convenir d'attribuer à un événement singulier la probabilité de l'événement générique qu'il réalise, c'est-à-dire la fréquence relative de cet événement générique dans une certaine suite. Toutefois, la mise en oeuvre de cette convention est problématique. En premier lieu, un événement singulier donné réalise de nombreuses propriétés, et donc de nombreux événements génériques, et il n'est pas clair duquel de ces événements il doit être considéré comme une réalisation. En second lieu, même si cet événement générique est identifié, il reste à déterminer relativement à quelle suite les probabilités doivent être définies. En d'autres termes, il reste à déterminer quelle est la suite dans laquelle les fréquences relatives sont les probabilités ; c'est le problème classique de la classe de référence. Surtout, même si les deux difficultés dont nous venons de faire état peuvent être surmontées, l'artifice que nous envisageons ne permet d'obtenir qu'un ersatz de la notion de probabilité d'un événement singulier. Popper le dit en termes linguistiques :

> ... un *énoncé de probabilité singulier* [c'est-à-dire un énoncé relatif à la probabilité d'un événement singulier] ... n'est singulier que sur le plan grammatical.[39]

[38] Ou un collectif. Distinguer entre les deux notions n'est pas nécessaire pour dire ce que nous avons à dire du fréquentisme dans ce travail ; nous les utilisons donc de manière indifférenciée.

[39] Popper (1983) p. 300.

Dans ces conditions, à celui qui d'une part considère que les probabilités sont objectives et d'autre part veut pouvoir penser des probabilités véritablement singulières, il reste à introduire une nouvelle théorie des probabilités. C'est précisément ce que fait Popper dans le milieu des années 1950.

Une autre justification du propensionnisme. Maintenant, il existe une justification du propensionnisme sensiblement différente de celle que nous venons de donner. Selon cette justification, l'apport du propensionnisme à la philosophie des probabilités consisterait dans le fait de considérer les probabilités comme des propriétés d'ensembles de conditions physiques là où le fréquentisme en fait des propriétés de collectifs.

A chaque fois qu'il présente cette vision des choses[40], Popper utilise l'exemple suivant :

> Supposons que nous avons un dé pipé et que nous nous sommes convaincus, à l'issue de longues suites d'expériences, que la probabilité d'obtenir un six avec ce dé pipé est très proche de 1/4. Considérez maintenant une suite, disons b, de lancers de ce dé pipé, mais qui inclut quelques lancers (deux, ou peut-être trois) d'un dé homogène et symétrique.[41]

Dans une telle situation, analyse Popper, le fréquentiste ne peut attribuer la valeur attendue 1/6 à la probabilité d'obtenir un six à l'occasion d'un lancer du dé équilibré que si :

1. il introduit une suite infinie ou très longue – en tout cas virtuelle – de lancers du dé équilibré ;
2. il considère que la probabilité d'obtenir un six à l'occasion d'un lancer du dé équilibré est la fréquence relative des six dans cette suite – et non dans la suite de départ, à laquelle les lancers du dé équilibré appartiennent pourtant.

Pour le dire autrement, le fréquentiste ne peut rendre compte de l'intuition selon laquelle la probabilité d'obtenir un six avec le dé équilibré est 1/6 qu'à la condition de préciser la notion de collectif dans un sens tel que la suite initiale n'est pas un collectif et la nouvelle suite en est un :

> ... il dira désormais qu'une suite admissible d'événements (suite de référence, ou encore "collectif") doit toujours consister en une répétition des mêmes conditions. Ou plus généralement, que les suites

[40] En particulier : Popper (1959) pp. 31 et suivantes, et Popper (1983) pp. 366 et suivantes.

[41] Popper (1959) p. 31.

> admissibles doivent être des suites virtuelles ou réelles qui sont caractérisées par un ensemble de conditions génératrices - ensemble dont la réalisation répétée produit les éléments d'une suite d'événements indépendants.[42]

Or, préciser en ce sens la notion de collectif reviendrait exactement à abandonner la théorie fréquentiste au profit de la théorie propensionniste :

> ...si nous examinons de plus près cette modification en apparence légère, nous découvrons qu'elle a pour effet de nous faire passer de l'interprétation fréquentiste à l'interprétation propensionniste.[43]

Assez clairement, cette seconde justification du propensionnisme est sensiblement plus favorable que la précédente au propensionnisme de long terme. Alors que la justification présentée dans le dernier paragraphe faisait du cas singulier la raison d'être du propensionnisme, celle dont nous venons de faire état conduit à considérer que le propensionnisme de long terme est le plus fidèle aux motivations initiales de Popper. Ainsi n'est-ce pas un hasard si l'argument dont nous venons de faire état est généralement introduit très tôt dans les défenses du propensionnisme de long terme.[44]

Néanmoins, un examen plus attentif révèle que la situation du propensionnisme de cas singuliers relativement à la justification que venons de présenter n'est pas l'analogue de la situation du propensionnisme de long terme relativement à la justification précédente. En effet, la première justification n'est d'aucune pertinence relativement au propensionnisme de long terme : si l'intérêt du propensionnisme est qu'il permet de penser des probabilités objectives d'événements singuliers, alors le propensionnisme de long terme n'a pas lieu d'être. De l'autre côté, si l'intérêt du propensionnisme est de mettre l'accent sur les conditions physiques qui engendrent les événements, alors le propensionnisme de cas singuliers reste une option ouverte : avant d'engendrer des fréquences relatives, un ensemble de conditions physiques donné engendre des événements singuliers.

Deux critiques du propensionnisme de long terme. A ce point, il nous semble qu'il reste au propensionniste de long terme une seule possibilité, qui consiste à 1) considérer que l'introduction du propensionnisme aux côtés du fréquentisme est motivée seulement par l'accent qu'il met sur les conditions physiques d'engendrement des événements et 2) soutenir que cet accent est mis à moindres frais par le propensionnisme de long terme que par

[42]Popper (1983) p. 368.
[43]Popper (1983) pp. 368–369.
[44]A titre d'illustration, voir Gillies (2000b).

le propensionnisme de cas singuliers. Le fait décisif serait donc que le propensionnisme de long terme diffère moins du fréquentisme que n'en diffère le propensionnisme de cas singuliers.

C'est ici que nous adressons au propensionnisme de long terme notre première critique. En effet, l'écart théorique entre fréquentisme et propensionnisme de long terme n'est pas seulement moindre que l'écart théorique entre fréquentisme et propensionnisme de cas singulier ; il est également fort mince. Ainsi, nous critiquons que le propensionnisme de long terme soit autre chose qu'un fréquentisme bien compris. C'est ce que suggère notre dernier paragraphe ; c'est ce que confirme l'analyse qui suit. Si les propensions tendent à réaliser des fréquences relatives proches des probabilités, la seule différence qui subsiste entre fréquentisme bien compris et propensionnisme de long terme consiste dans l'hypothèse des propensions derrière les fréquences relatives. Or cette différence ne concerne pas ce que sont les probabilités finalement. Kyburg le dit dans les termes suivants :

> selon la conception de long terme, c'est la suite entière d'essais, qu'elle soit actuelle ou hypothétique, qui incarne la probabilité. Mais là, dans une suite, "propension" ne revient pas à autre chose que fréquence relative de long terme – un concept assez commun (*a pedestrian enough concept*).[45]

Selon cette analyse, rien ne justifie d'introduire le propensionnisme de long terme à côté d'un fréquentisme correctement compris, du type de celui de Kolmogorov par exemple[46].

La seconde de nos critiques au propensionnisme de long terme est en un sens plus subjective : nous ne comprenons pas exactement ce que peut être une propension à produire des fréquences relatives. Plus exactement, le concept de propension à produire des fréquences relatives nous semble difficilement intelligible dès lors qu'on prend au sérieux l'idée des propensions comme entités métaphysiques (au sens du paragraphe 5.1.1.4) actives qui tendent à la réalisation d'événements observables. Eagle explicite la difficulté :

> Si la propension est effectivement inactive dans le cas singulier, mais seulement dans le long terme, la propension ne peut être identifiée avec aucune disposition locale de chaque essai. Soit la propension est elle-même une fusion des dispositions à chaque essai ; soit le type d'essai [pertinent] doit perdurer dans le temps, la propension étant pleinement présente à chaque instant. Chacune de ces deux options semble appeler

[45] Kyburg (2002) p. 15.

[46] Sur ce point, voir par exemple Gillies (2000b) p. 811.

la controverse.[47]

Ainsi la notion de propension à réaliser des fréquences relatives pose-t-elle un problème d'intelligibilité dès lors qu'on l'envisage comme distincte de la notion de propension à réaliser des événements singuliers. Il nous semble qu'elle pose également un problème de définition, relatif à la longueur minimale d'une suite dans laquelle les fréquences relatives résultent de l'activité de propensions de long terme.

Dans les paragraphes 5.1.2.2 et 5.1.2.3, nous avons répondu aux deux critiques les plus fréquentes du propensionnisme de cas singuliers ; dans le paragraphe 5.1.2.4, nous avons donné des raisons positives de préférer le propensionnisme de cas singuliers au propensionnisme de long terme. En conformité avec ces raisons, nous nous rangeons au propensionnisme de cas singuliers. C'est cette théorie que nous désignerons maintenant au moyen du terme générique de « propensionnisme ». Ainsi que nous l'avons annoncé au début du présent chapitre, un aspect de ce travail consiste à mettre au jour les positions philosophiques dont le propensionnisme est solidaire. Toutefois avant cela – et afin d'achever notre caractérisation du propensionnisme –, nous consacrons une sous-section à la question de savoir si le propensionnisme de cas singuliers est une interprétation du calcul des probabilités.

5.1.3 Le propensionnisme, une interprétation du calcul des probabilités ?

De même que celle qui a été menée dans le paragraphe 5.1.1.3, la discussion qui commence ne prend pas en compte les probabilités conditionnelles. Pour le dire autrement, la question que nous posons ici est celle de savoir si le propensionnisme de cas singuliers tel que nous l'avons présenté dans ce qui précède peut être considéré comme une interprétation du calcul des probabilités absolues. Les questions de savoir ce que sont les propensions conditionnelles et surtout si elles sont des probabilités conditionnelles seront abordées plus loin dans le prochain chapitre.

5.1.3.1 Interpréter les probabilités

Ainsi que le remarquent de nombreux auteurs[48], l'idée d'*interpréter* le calcul des probabilités est historiquement située. Plus précisément, elle émerge dans la foulée de l'axiomatisation du calcul des probabilités par Kolmogorov

[47] Eagle (2004) p. 399.

[48] En particulier Gillies (Gillies (2000b) section 4 par exemple) et Humphreys (Humphreys (1985) section IV par exemple).

dans *Foundations of the Theory of Probability* et corrélativement de l'idée selon laquelle le calcul des probabilités serait une extension de la logique classique[49]. Le terme « probabilité » est alors considéré comme un terme primitif à interpréter. De façon plus générale, interpréter les probabilités s'entend alors au sens suivant : produire un modèle de cette théorie formelle qu'est la théorie de probabilités. En vue de déterminer si le propensionnisme de cas singuliers est une interprétation du calcul des probabilités en ce sens, il convient donc de rappeler quels sont les axiomes de Kolmogorov pour les probabilités.

Dans la présentation de Kolmogorov, une fonction de probabilité est toujours définie relativement à un ensemble probabilisable (Ω, $\mathbf{O}$). Ici, Ω est un ensemble non vide et $\mathbf{O}$ est une σ-algèbre de parties de Ω – c'est-à-dire un ensemble de sous-ensembles de Ω auquel Ω appartient et qui est clos par réunion et par complémentation. Etant donné, maintenant, un tel couple (Ω, $\mathbf{O}$), une fonction P de $\mathbf{O}$ dans l'ensemble des nombres réels est une fonction de probabilités si :

1. pour tout E de $\mathbf{O}$, $\mathrm{P(E)} \geq 0$
2. $\mathrm{P}(\Omega) = 1$
3. si $(\mathrm{E,F}) \in \mathbf{O}^2$ et $\mathrm{E} \cap \mathrm{F} = \emptyset$, alors $\mathrm{P(E} \cup \mathrm{F)} = \mathrm{P(E)} + \mathrm{P(F)}$
4. si $(\mathrm{O}_i)_{i\in\mathbb{N}}$ est une famille d'éléments de $\mathbf{O}$ deux à deux disjoints, alors $\mathrm{P}(\bigcup_{i=1}^{+\infty} \mathrm{O}_i) = \sum_{i=1}^{+\infty} \mathrm{P}(\mathrm{O}_i)$.

Si P est une fonction de probabilités, alors le triplet (Ω, $\mathbf{O}$, P) est un espace de probabilités. Pour des raisons qui sont largement de commodité, nous ne tiendrons pas compte dans la suite du dernier axiome – dit d'« additivité dénombrable ».

Si l'on s'en tient strictement à la notion modèle-théorique d'interprétation, interpréter le calcul des probabilités est définir :

– un ensemble non vide ω,
– une σ-algèbre $\mathbf{o}$ de parties de ω et
– une fonction p de $\mathbf{o}$ dans $\mathbb{R}$

tels que les formules 1. à 3. sont satisfaites. Or, ainsi explicitée, la notion modèle-théorique apparaît clairement trop stricte. Pour le comprendre, considérons à titre d'illustration la théorie subjectiviste telle qu'elle est développé par de Finetti. On s'accorde à considérer qu'il s'agit d'une interprétation du calcul des probabilités. Or, nous serions bien en peine d'identifier ce que seraient ω et surtout $\mathbf{o}$ dans ce cadre. En effet, les probabilités subjectives ne se présentent pas comme des probabilités d'*ensembles*, mais

[49]Humphreys rappelle que cette conception a été « adoptée par Bolzano, Boole, Venn, Lukasiewicz, Reichenbach, Carnap et Popper » (Humphreys (1985) p. 568).

comme des probabilités d'*événements*. Plus généralement, et ainsi que le remarque Popper :

> dans l'approche de Kolmogorov, on suppose que les objets a et b dans p(a,b) sont des ensembles (ou des agrégats). Mais cette supposition n'est pas partagée par toutes les interprétations.[50]

Il en déduit qu'il est nécessaire de proposer des « axiomes d'un caractère seulement métrique »[51]. Avant cela, nous en déduisons que « interprétation » a un sens plus faible dans l'expression « interprétation du calcul des probabilités » qu'en théorie des modèles.

Positivement, il semble à ce point qu'interpréter le calcul des probabilités est définir :

- un ensemble **o'** qui est clos par union et par complémentation ;
- un élement ω' de **o'**
- une fonction p' de **o'** dans $\mathbb{R}$

tels que les formules 1. à 3. satisfaites. A titre d'illustration de ce qu'il faut entendre par là, revenons à de Finetti. Nous avons vu qu'il interprète les probabilités comme des probabilités d'événements. Autrement dit, **o'** est l'ensemble des événements. Dans ce cas particulier, 3. est satisfait si et seulement si :

> pour tout couple (E,F) d'événements incompatibles,
> p'(E $\vee$ F) = p'(E) + p'(F),
> où E $\vee$ F est l'événement qui advient si et seulement si E advient ou F advient.

Selon le théorème de Ramsey – de Finetti, le subjectivisme constitue bien une interprétation du calcul des probabilités au sens corrigé que nous venons de définir. Un résultat analogue est encore plus simple à obtenir dans le cas du fréquentisme. Ce qu'il nous faut déterminer maintenant est ce qu'il en est du propensionnisme.

5.1.3.2 Difficultés pour le propensionnisme

De même que de Finetti, le propensionniste propose de considérer que ω' est un ensemble d'événements. La condition 3. se formule alors comme précédemment. La notion de probabilité, quant à elle, est toujours relative à un ensemble de conditions physiques. Celui-ci étant donné, la probabilité p'(E) d'un événement est la mesure de la propension qui tend à réaliser E.

La question qui se pose d'abord – c'est-à-dire en relation directe avec le premier des axiomes de Kolmogorov – est celle de l'intervalle dans lequel

[50] Popper (1959) p. 40.
[51] Popper (1959) p. 40.

cette mesure prend ses valeurs. A cette question, Popper apporte une réponse conventionnaliste :

> Les *probabilités mathématiques* [...] prennent leurs valeurs numériques de 0 à 1[52]

ou :

> la propension maximale correspond au cas particulier d'une force classique en acte [...]. Lorsque la propension est inférieure à A, on peut interpréter cela comme dénotant l'existence de forces concurrentes [...].[53]

Notons que la nécessité de recourir à une convention ne tient pas au fait que, sous l'interprétation propensionniste, les probabilités sont des *mesures*, et que définir une mesure implique de définir une unité et/ou une intervalle de valeurs. Pour nous en convaincre, nous rappellerons d'une part que les degrés de croyance des subjectivistes sont eux aussi des mesures – de l'intensité des croyances – et d'autre part qu'on *montre* qu'ils doivent prendre leur valeur dans [0 ; 1]. Maintenant, si on peut le montrer, c'est que le subjectivisme comporte une théorie de la mesure des degrés de croyances, et donc *comporte une théorie de la mesure des probabilités*. Plus explicitement, la thèse selon laquelle les degrés de croyance sont mesurés dans les situations de pari fait partie intégrante de la position subjectiviste en philosophie des probabilités. A l'inverse, le propensionnisme ne propose pas de théorie de la mesure des propensions.

De ce que le propensionnisme ne comporte pas une théorie de la mesure des propensions, et donc de l'évaluation des probabilités, il découle que la question de savoir si le propensionnisme de cas singuliers est une interprétation du calcul des probabilités n'a pas de réponse empirique. Cela n'implique pas que le propensionnisme *n'est pas* une interprétation du calcul des probabilités. Aussi, la question de savoir si le propensionnisme est une interprétation du calcul des probabilités reste, à ce point, ouverte. Le prochain paragraphe est consacré à envisager deux arguments qui peuvent être avancés en faveur de la thèse selon laquelle il en est une.

5.1.3.3 Deux arguments en faveur de la thèse selon laquelle le propensionnisme de cas singuliers est une interprétation du calcul des probabilités

Retour sur les critères d'adéquation. Le premier des deux arguments que nous développons dans le paragraphe qui commence stipule que la notion

[52]Popper (1990) p. 33.
[53]Popper (1990) p. 34.

d'interprétation du calcul des probabilités telle que nous l'avons thématisée n'est pas satisfaisante. Cette thèse est défendue en particulier dans la dernière section de Humphreys (1985). Dans cette section, Humphreys fait fonds sur l'analyse selon laquelle les propensions conditionnelles ne sont pas des probabilités conditionnelles. Nous avons dit déjà que cette question n'est pas abordée dans le présent chapitre. Cela, toutefois, n'implique pas qu'il nous soit impossible de nous appuyer ici sur les arguments de Humphreys. Positivement, Humphreys donne, contre la notion d'interprétation des probabilités que nous venons de développer, deux arguments indépendants de la considération des propensions conditionnelles. Ce sont ces arguments qu'il convient que nous exposions maintenant.

Le premier des deux arguments que nous identifions dans Humphreys (1985) consiste d'abord à faire valoir que la notion d'interprétation que nous venons de développer découle du projet, historiquement ancré, de faire des probabilités une branche de la logique. Il consiste ensuite à soutenir que ce projet, s'il « a eu un effet clarificateur énorme en ce qui concerne les investigations sur les probabilités »[54], semble avoir fait long feu. Plus précisément, la diversité de mieux en mieux aperçue des objets de la théorie des probabilités impose de renoncer au projet d'une axiomatisation unique. Corrélativement, l'exigence en vertu de laquelle toute théorie des probabilités devrait assurer la satisfaction des axiomes de Kolmogorov – c'est-à-dire : devrait être admissible – doit être abandonnée. Dans les termes de Humphreys : « Il est temps, je pense, d'abandonner le critère d'admissibilité »[55].

Le second des arguments que nous identifions dans la section IV. de Humphreys (1985) vient à l'appui du premier. Il consiste à faire remarquer que le critère d'admissibilité n'est pas ce qui nous permet de reconnaître effectivement qu'une théorie est une interprétation du calcul des probabilités. En effet, les interprétations reconnues du calcul des probabilités ne satisfont pas ce critère :

> Ajoutez à cela que les fréquences relatives dans des suites finies violent l'axiome d'additivité dénombrable[56] et que leurs espaces de probabilités ne sont pas des σ-algèbres à moins d'imposer des contraintes supplémentaires, que les degrés de croyance rationnels, sous certaines conceptions, ne sont pas dénombrablement additifs et qu'on ne peut pas raisonnablement exiger qu'ils le soient, et qu'on peut douter sérieusement que la théorie des probabilités traditionnelle soit adaptée

[54]Humphreys (1985) p. 570.
[55]Humphreys (1985) p. 569.
[56]Il s'agit de l'axiome 4. ci-dessus.

à la mécanique quantique.[57]

Mais si l'admissibilité, c'est-à-dire le fait d'assurer que la satisfaction des axiomes de Kolmogorov, n'est pas un critère satisfaisant, à quoi reconnaît-on qu'une théorie est une interprétation du calcul des probabilités ? Pour répondre à cette question, on peut prendre appui sur une analyse de Salmon. Dans Salmon (1966), il propose trois « critères d'adéquation » pour les interprétations du calcul des probabilités :

- l'admissibilité – dont nous venons de parler, c'est-à-dire le fait d'assurer la satisfaction des axiomes du calcul des probabilités :

  > Nous disons qu'une interprétation d'un système formel est admissible si la signification assignée aux termes primitifs par cette interprétation transforme les axiomes formels en affirmations vraies[58] ;

- la possibilité d'évaluer les probabilités (*ascertainability*), c'est-à-dire le fait que « au moins en principe, nous puissions évaluer les probabilités »[59] ;
- l'applicabilité, c'est-à-dire la capacité à rendre compte des applications des probabilités :

  > La force de ce critère est le mieux exprimée par le fameux aphorisme de l'évêque Butler : "La probabilité est le guide même de la vie" (*probability is the very guide of life*).[60]

Maintenant, nous venons de voir que l'admissibilité n'est pas nécessaire pour parler d'interprétation des probabilités. Dans ces conditions, les trois critères doivent être envisagés comme des descriptions d'échelles relativement auxquelles une théorie des probabilités donnée peut être plus ou moins satisfaisante.

Dans ce nouveau cadre d'évaluation du statut des théories des probabilités, le propensionnisme a des avantages à faire valoir du côté de l'applicabilité. Hájek (2007) identifie des composantes du critère d'applicabilité. L'examen de ces composantes suggère que la théorie propensionniste des probabilités a les avantages suivants :

- elle permet de penser des probabilités non triviales ;
- elle distingue (contrairement aux interprétations fréquentistes) entre les probabilités et les fréquences relatives ;
- elle rend compte de ceci que les événements les plus probables se produisent le plus fréquemment. Cette thèse n'est intelligible que si

[57] Humphreys (1985) pp. 569–570.
[58] Salmon (1966) p.63.
[59] Salmon (1966) p. 64.
[60] Salmon (1966) p. 64.

« événements » s'entend au sens générique. Mais, alors, il convient d'expliquer pourquoi elle est vraie. Le raisonnement est ici le suivant. D'une part, les propensions sont attachées aux ensembles de conditions physiques ; d'autre part, une expérience aléatoire est définie par – et donc, finalement, comme – un ensemble de conditions physiques. Dans ces conditions, les mêmes propensions sont à l'oeuvre à chaque répétition d'un ensemble de phénomènes aléatoires. Il existe alors, à chaque répétition de l'expérience, de fortes propensions à produire les instanciations singulières d'un même événement générique très probable. Ces instanciations se produisent donc plus souvent que les instanciations d'événements génériques peu probables ;

– elle est la théorie qui rend le meilleur compte des utilisations des probabilités en mécanique quantique.

Relativement à la troisième des échelles d'évaluation des théories des probabilités, le propensionnisme ne peut donc pas être considéré comme une théorie nulle. Avant de conclure sur le point du statut du propensionnisme, nous rendons compte d'un dernier argument dont le propensionniste dispose dans le débat qui nous intéresse.

L'argument de Lewis. L'argument que nous présentons est introduit dans Lewis (1980). Il vise à montrer que les propensions satisfont bien, finalement, les axiomes 1. à 3. de Kolmogorov. En d'autres termes, il s'agit de proposer un argument (nécessairement non empirique) en faveur de la thèse selon laquelle le propensionnisme se situe au même niveau que le subjectivisme sur l'échelle d'évaluation des théories des probabilités qui correspond au critère d'admissibilité. Cette conclusion est tirée de deux propositions :

1. les propensions contraignent les degrés de croyance rationnelle selon les voies explicitées par le Principe Principal de Lewis :

   Proposition 5.1 (Principe Principal (Lewis, 1980)) *Si P est la fonction subjective de probabilités d'un agent rationnel et ch est la fonction objective de probabilités singulières (*chance fonction*),*
   alors $P(A|ch(A) = x)) = x$,
   pour tout A et tout x tels que $P(ch(A) = x) \neq 0$.

2. les degrés de croyance rationnelle satisfont prouvablement les axiomes 1. à 3. de Kolmogorov.[61]

L'idée est donc de faire découler la satisfaction des axiomes de Kolmogorov dans le cadre du propensionnisme de leur satisfaction dans le cadre du sujectivisme :

[61] Il s'agit du résultat de Ramsey – de Finetti (Ramsey (1926), de Finetti (1937)).

> Quoi que ce soit qui vient d'une distribution de probabilités par conditionalisation, est une distribution de probabilités. En conséquence, une distribution objective de probabilités singulières est une distribution de probabilités.[62]

Ainsi que le remarque Hájek dans Hájek (2007), l'argument que nous venons de présenter trouve à s'appuyer sur la thèse selon laquelle les termes théoriques sont définis par le rôle qu'ils jouent dans les théories. Cette approche semble particulièrement adaptée à la définition des propensions conçues comme objets de nature dispositionnelle. Les propriétés dispositionnelles, en effet, se caractérisent d'abord en termes fonctionnels. Néanmoins, l'argument n'est concluant que si le Principe Principal est accepté dans la forme que lui donne initialement Lewis, et qui a été critiquée ensuite. Dans ces conditions, nous considérons qu'il n'est pas concluant mais mettrons au crédit du propensionnisme le fait qu'il doit exister une forme de rapport entre les propensions et les degrés de croyance rationnelle.

En conclusion, nous avons montré que la question de savoir si le propensionnisme est une interprétation des probabilités au sens strict du critère d'admissibilité est une question qui n'a pas de réponse empirique. Cela n'implique pas que le propensionnisme doive être rejeté, ni même qu'il n'ait pas d'intérêt comme théorie des probabilités. A partir des critères d'adéquation proposés dans Salmon (1966), nous avons fait apparaître plusieurs caractéristiques intéressantes du propensionnisme. Puis, nous avons présenté l'argument lewisien en faveur de la thèse selon laquelle les probabilités comprises dans les termes de la théorie propensionniste satisfont les trois premiers axiomes de Kolmogorov. Dans ces conditions, nous acceptons pour la suite de notre travail l'hypothèse selon laquelle le propensionnisme de cas singuliers est une interprétation du calcul des probabilités. Il nous resterait à déterminer si le propensionnisme peut également être considéré comme une interprétation du calcul des probabilités étendu aux probabilités conditionnelles. Cette question est traitée dans le prochain chapitre. Avant d'en arriver là, et conformément à ce que nous avons annoncé, nous nous arrêtons le temps d'une section aux corrélats philosophiques de la théorie propensionniste des probabilités absolues.

[62] Lewis (1980) p. 277.

5.2 Corrélats philosophiques du propensionnisme

Nous avons caractérisé les propensions comme des entités physiques 1. inobservables, qui 2. tendent à réaliser des événements – et donc ont quelque chose comme une puissance causale – et 3. sont des propriétés dispositionnelles. La conjonction de ces trois éléments a conduit de nombreux auteurs à considérer le propensionnisme comme une théorie métaphysique – dans tous les sens possibles de ce qualificatif. Elle indique en tout cas clairement que cette théorie a des corrélats philosophiques importants. Ce sont eux que nous nous proposons d'expliciter dans la section qui commence. Cette explicitation comporte trois temps, correspondant à trois domaines importants de la philosophie : l'ontologie, l'épistémologie et la métaphysique. Elle s'appuie largement sur des comparaisons entre le propensionnisme et les autres théories des probabilités.

5.2.1 Ontologie propensionniste

5.2.1.1 Le propensionnisme dans la tradition ontologique

Nous entendons par « ontologie » une théorie de ce qui est. L'histoire de la philosophie conduit à en distinguer deux grands types, qui s'opposent sur la question de l'existence d'autre chose que des instanciations locales de propriétés. En termes plus contemporains et plus précis, le point de désaccord concerne l'existence d'entités irréductibles – en un sens à préciser – à des points d'espace-temps occupés ou non.

La réponse « non » à cette question donne lieu à une première tradition, allant de Démocrite à Hume, puis à Quine et Lewis. La formulation la plus frappante de ce minimalisme ontologique est sans doute proposée par Lewis :

> Tout ce qu'il y a au monde est une vaste mosaïque de faits locaux particuliers (*local matters of particular facts*) : juste une petite chose et puis une autre.[63]

La seconde tradition ontologique, à l'inverse, est prête à admettre autre chose que des points d'espace-temps. Les entités contribuent d'ailleurs le plus souvent à expliquer la distribution spatio-temporelle des propriétés locales, qui perd son statut de caractéristique ultime du monde. Un des premiers représentants de cette tradition est à nos yeux Aristote ; de manière générale, les philosophes classiques, de Descartes à Leibniz, en relèvent. Le propension-

[63] Lewis (1986) p. (ix).

nisme contribue à y faire figurer Popper. D'une part, en effet, les propensions existent pour lui au moins au même titre que ce qui est observable :

> On suppose que les propensions ne sont pas de simples possibilités, mais qu'elles ont une réalité physique.[64]

D'autre part, on peut considérer qu'elles expliquent les événements qu'elles contribuent à produire. La relation qui unit les propensions à la distribution spatio-temporelle des propriétés locales est donc telle que les premières sont ontologiquement irréductibles aux points d'espace-temps qui supportent la seconde. Le propensionnisme impose donc une ontologie qui accepte des entités autres que les points d'espace-temps.

Nous nous proposons d'étudier en quoi cette position contribue à le distinguer parmi les théories des probabilités. Pour cela, nous nous intéresserons successivement aux deux principales théories concurrentes du propensionnisme, sa cousine la plus proche – la théorie fréquentiste – d'abord, son ennemie la plus explicite – la théorie subjectiviste – ensuite. Quoique fort différentes, ces deux théories ont en commun d'être développées dans les années 1930 et souvent présentées comme des théories opérationalistes des probabilités. A proprement parler, l'opérationalisme est la thèse selon laquelle :

> chaque concept nouveau introduit en physique doit être défini de façon opérationnelle en termes de procédures expérimentales, et de concepts déjà définis.[65]

Plus généralement, tous les termes théoriques sont définis par l'énonciation des procédures expérimentales permettant de leur assigner une grandeur – de les mesurer.

L'opérationalisme est une option épistémologique. Nous abordons la question de l'opérationalisme dans cette sous-section dans la mesure où celui-ci constitue une position réductionniste dont les conséquences ontologiques sont clairement déflationnistes. En effet, si les termes théoriques se résolvent complètement dans la description de procédures effectives, alors ceux-ci sont ontologiquement réductibles à celles-là. L'opérationalisme conduit donc à répondre négativement à la question de l'existence d'entités irréductibles aux points d'espace-temps. Nous nous proposons donc d'étudier en quoi les définitions fréquentiste et subjectiviste du concept de probabilité sont opérationalistes. Ici, notre analyse s'appuiera largement sur Gillies (2000a).

[64] Popper (1990) p. 33.
[65] Gillies (1972) p. 6.

5.2.1.2 Ontologie fréquentiste

Dans sa version la plus attractive, le fréquentisme définit la probabilité comme une fréquence-limite. Plus précisément, la probabilité d'un attribut dans une suite d'événements – c'est-à-dire dans une suite d'événements uniformes en un certain sens et différant par des attributs observables – est la limite de la fréquence relative de cet attribut quand la longueur de la suite tend vers l'infini. Gillies s'attache à montrer que cette définition est opérationaliste. Il cite alors Von Mises :

> la fréquence relative de la répétition est la "mesure" de la probabilité, exactement de la même façon que la longueur de la colonne de mercure est la "mesure" de la température.[66]

Cette analogie semble bien indiquer en quoi la définition fréquentiste de la probabilité est opérationaliste.

Pourtant, elle suscite d'emblée la réticence : Von Mises ne définit pas la probabilité comme fréquence relative dans une suite empirique finie – même très longue –, mais comme fréquence-limite dans un suite infinie, et donc nécessairement idéale. La fréquence relative ne mesure donc pas la probabilité au sens effectif où la colonne de mercure mesure la température. Gillies fait droit à cette objection :

> La définition de la probabilité comme fréquence-limite est supposée être une définition opérationaliste d'un terme théorique (probabilité) en termes d'un concept observable (*observable concept*) (fréquence). Cependant, on pourrait arguer qu'elle échoue à proposer une connexion entre observation et théorie à cause de l'utilisation de limites dans une suite infinie.[67]

On pourrait proposer une connexion entre observation et théorie en revenant à la définition weierstrassienne de la limite dans une suite. Selon cette définition, pour que p soit la limite de la fréquence de l'attribut A dans la suite S, il faut, pour tout réel ϵ, pouvoir exhiber un rang n à partir duquel la fréquence relative de A dans S diffère de p par ϵ au plus.[68] L'idée

[66] Von Mises, préface à la troisième édition allemande de *Probability, Statistics and Truth*, 1950. Cité dans Gillies (2000a) p. 100.

[67] Gillies (2000a) p. 101.

[68] Ce que nous envisageons ici est une opérationalisation du concept de probabilité *tel qu'il est compris par le fréquentisme*, c'est-à-dire comme fréquence-limite. En faisant cela, nous ne prétendons qu'il n'y a pas de problème à considérer que les probabilités sont des limites de fréquences. En particulier, nous ne prétendons pas que le fréquentisme est capable de rendre compte de ceci – que nous avons vu – que la convergence des limites de fréquences est seulement une convergence en probabilité.

d'utiliser cette définition pour connecter les fréquences-limites aux fréquences observées achoppe sur deux difficultés :

- la question à laquelle elle permet de répondre n'est pas : "quelle est la probabilité de A dans S ?", mais : "la probabilité de A dans S est-elle p ?" ;
- cette question n'est pas décidable (mais seulement semi-décidable) ; la procédure associée n'est donc pas effective.

Dans ces conditions, il semble que revenir à la définition des suites ne permet de pas de rendre opérationaliste la définition fréquentiste des probabilités.

Plutôt que de s'engager dans la voie de l'établissement d'une connexion entre observation et théorie, Gillies suggère qu'il est possible de se passer d'une telle connexion. Reprenant un argument de Von Mises, il fait valoir que des concepts de physique mathématique bien admis reposent par définition sur de « telles représentations du fini par l'infini »[69]. L'argument de Von Mises que reprend Gillies s'inscrit dans un projet général visant à faire de la théorie des probabilités une science mathématique épistémologiquement homogène à la mécanique. Mais il ne montre pas que le fréquentisme peut bien, finalement, être considéré comme une théorie opérationaliste.

Au contraire, il est apparu au cours de sa discussion que la notion fréquentiste de probabilité n'est définie que relativement à des suites infinies idéales. Dès lors, ces suites existent et, surtout, elles existent en un sens différent de celui où existent les suites empiriques dont ils procèdent par idéalisation. Elles sont ontologiquement irréductibles aux fréquences relatives observables. Autrement dit, la définition fréquentiste des probabilités suppose une ontologie ouverte à des entités irréductibles aux points d'espace-temps. Le fréquentisme et le propensionnisme ne se distinguent donc pas par des réponses différentes à la question ontologique fondamentale de l'existence de telles entités. Pour distinguer ces deux théories relativement à l'ontologie qu'elles supposent, il faut recourir à une distinction plus fine. Celle-ci pourrait être relative au type des entités inobservables admises. Nous pouvons en effet opposer le caractère naturel des propensions au caractère idéal des suites infinies. La notion d'idéalité mériterait d'être discutée pour elle-même. Il semble néanmoins clair que les suites infinies n'appartiennent pas à la réalité physique ; c'est ce que nous retiendrons.

C'est donc la nature physique des propensions qui fait l'originalité ontologique du propensionnisme en regard des théories fréquentistes des probabilités. Cette originalité est seulement secondaire – au sens où elle ne repose pas sur une réponse inédite à la question ontologique fondamentale de l'existence exclusive des points d'espace-temps. Nous nous proposons de montrer

[69] Gillies (2000a) p. 101.

maintenant qu'il en va autrement si nous comparons l'ontologie propensionniste à l'ontologie à laquelle les subjectivistes en philosophie des probabilités sont engagés à souscrire.

5.2.1.3 Ontologie subjectiviste

Nous avons pris pour point de départ de notre analyse la réputation de théorie opérationaliste commune aux théories fréquentiste et subjectiviste du calcul des probabilités. Dans le cas du fréquentisme, cette réputation s'est avérée mal fondée. Qu'en est-il pour le subjectivisme ? Il nous faut l'examiner ici, dans la mesure où, rappelons-le, l'opérationalisme conduirait à répondre positivement à la question de l'existence exclusive des points d'espace-temps.

Probabilités subjectives et opérationalisme. Les subjectivistes définissent la probabilité comme un degré de croyance rationnelle susceptible d'être mesuré – c'est ce qui nous importe ici – dans les situations de pari. Ainsi de Finetti soutient-il que, pour donner une « définition quantitative » de la notion de probabilité,

> il s'agit simplement de préciser mathématiquement l'idée banale et évidente que le degré de probabilité attribué par un individu à un événement donné est révélé par les conditions dans lesquelles il serait disposé à parier sur cet événement.[70]

Ce qui sera fait quelques lignes plus bas. Nous ne reprenons pas cette définition qui, en tant que telle, ne présente guère d'intérêt pour nous ici. Il nous suffira d'avoir indiqué que le concept de probabilité est susceptible d'une définition consistant dans la description d'une procédure permettant de mesurer des degrés de croyance. Il en découle en effet que le terme théorique « probabilité » est ontologiquement réductible à cette procédure et ne conduit donc pas à supposer l'existence d'autre chose que des points d'espace-temps.

Nous pourrions arrêter là notre analyse des présupposés ontologiques des théories subjectivistes des probabilités. Nous conclurions alors que ces théories proposent bien une définition opérationaliste du concept de probabilité, et qu'il en découle une différence ontologique fondamentale par rapport au propensionnisme. Nous aimerions toutefois suggérer que, en-deçà de la définition opérationaliste du concept de probabilité, le projet subjectiviste en philosophie des probabilités peut être tout entier conçu comme un projet ontologique visant à penser la probabilité sans recourir à des entités irréductibles aux points d'espace-temps.

[70] De Finetti (1937) p. 6.

Le projet subjectiviste comme projet ontologique. Pour le comprendre, revenons à de Finetti (1937). De Finetti en exprime clairement l'objet : il s'agit de montrer qu'une interprétation subjectiviste du calcul des probabilités est tenable. Pour cela, il doit se rendre capable « de ramener dans le cadre de la conception subjective et d'expliquer *même* les questions qui semblent la démentir »[71]. Ces questions épineuses sont au nombre de trois :

> Il y a trois objections essentielles : on doute que la conception subjective permette de définir la probabilité, de démontrer les lois logiques qui la régissent et enfin d'expliquer et justifier les applications qu'on en a fait aux problèmes les plus divers.[72]

De Finetti considère que la dernière de ces objections constitue « le problème le plus délicat »[73]. En effet, les utilisations quotidiennes des probabilités révèle souvent une concordance des opinions subjectives qui fait problème pour de Finetti. Il considère qu'elle l'engage à :

> *montrer qu'il y a des raisons psychologiques assez profondes pour rendre très naturelle la concordance exacte ou approchée qu'on observe entre les opinions des divers individus, mais qu'il n'y a pas de raisons rationnelles, positives, métaphysiques, qui puissent enlever à ce fait le caractère d'une simple concordance d'opinions subjectives.*[74]

Autrement dit, de Finetti refuse la solution consistant à rendre compte de la coïncidence des opinions subjectives en postulant l'existence d'une entité objective sur laquelle elles porteraient communément. Ce refus constitue une véritable prise de position : alors que la solution refusée est évidente, la solution choisie est conceptuellement moins économique. Elle impose en effet l'introduction de nouveaux concepts – notamment celui d'« événements équivalents » – et des développements techniques sinon très lourds, du moins non triviaux.

La prise de position que nous venons de mettre au jour est de nature ontologique. Il s'agit en effet de ne pas supposer d'entité dont l'existence porterait le poids d'une objectivité de la notion de probabilité. Surtout, elle est au fondement du projet subjectiviste en philosophie des probabilités : de Finetti (1937) est presque exclusivement consacré à la défendre ; « La probabilité n'existe pas » est le mot d'ordre *Theory of Probability*. Or, elle rend les probabilités ontologiquement réductibles aux estimations individuelles de probabilités singulières – c'est-à-dire, au final, à des points d'espace-temps.

[71] De Finetti (1937) p. 3.
[72] De Finetti (1937) p. 59.
[73] De Finetti (1937) p. 51.
[74] De Finetti (1937) p. 61. Les italiques sont dans le texte original.

Nous venons de montrer que le subjectivisme en philosophie des probabilités consiste, du point de vue ontologique, à définir le concept de probabilité sans outrepasser la conception minimaliste selon laquelle seuls existent les points d'espace-temps. L'engagement ontologique dont nous venons de faire état est relatif à la seule définition du concept de probabilité. Au premier abord, on voit mal comment il pourrait en être autrement : on n'attend pas d'une théorie des probabilités qu'elle engage à des thèses d'ontologie générale. Il nous semble pourtant que c'est le cas, ainsi que nous le montrons dans le prochain paragraphe.

De l'ontologie des probabilités à l'ontologie générale. Dans « La prévision », de Finetti indique les voies d'une généralisation des thèses ontologiques subjectivistes. Son argumentation est fondée sur la distinction de deux types de lois physiques : probabilistes d'une part, déterministes de l'autre. Si les probabilités n'ont de réalité que subjective, les lois probabilistes énoncent des régularités purement subjectives. A l'inverse, il est généralement admis que les lois déterministes énoncent des régularités objectives de la nature. De façon rhétorique, de Finetti demande : « N'y aurait-il pas là un abîme infranchissables séparant ces deux types de lois qui coexistent aujourd'hui en physique ? »[75]. Une telle situation serait rationnellement insatisfaisante. De Finetti en prend acte et propose un dépassement – le seul qui soit compatible avec sa position en philosophie des probabilités :

> Pour franchir cet abîme, le point de vue adopté jusqu'ici nous conduit tout naturellement à une solution qui est exactement l'opposée de celle qu'on envisage habituellement : au lieu d'étendre le caractère de réalité des lois classiques aux lois de probabilité, on peut essayer au contraire de faire participer ces mêmes lois classiques au caractère subjectif des lois statistiques.[76]

Le souci d'unité théorique conduit donc à ne plus concevoir de régularités que subjectives.

Cette thèse se comprend d'abord au plan métaphysique : le monde n'est pas objectivement structuré. Mais elle peut être aussi comprise au plan ontologique : s'il n'y a de régularités que subjectives, alors les lois de la nature n'existent pas. L'ontologie subjectiviste est donc fermée aux lois de la nature conçues comme entités objectives ; surtout, elle n'admet aucune forme de connexion nécessaire entre les entités qu'elle accepte. Cette dernière formulation permet de comprendre comment de Finetti en vient à considérer la notion de causalité, qui est bien celle d'une connexion nécessaire entre

[75]De Finetti (1937) p. 64.
[76]De Finetti (1937) p. 64.

événements. Il peut alors faire référence à Hume – et le passage de la notion de loi déterministe à celle de causalité a évidemment pour raison d'être d'autoriser cette référence :

> ... cette explication [celle qu'il vient de donner de la notion de loi déterministe] semble constituer la véritable traduction logique de la conception de "cause" préconisée par David Hume, que je considère comme le plus haut sommet qui ait été atteint par la philosophie.[77]

Or, la théorie de la causalité est essentielle à l'empirisme de Hume : elle fonde sa théorie empiriste de la connaissance, qui elle-même trouve son fondement et sa nécessité dans la thèse ontologique forte selon laquelle c'est aux seules perceptions – impressions (« les plus vives ») ou idées (« les plus faibles ») – que l'existence peut être attribuée.[78] Dans ces conditions, la référence à Hume suggère une extension à l'ontologie générale de la thèse d'abord limitée au domaine des probabilités, et selon laquelle les points d'espace-temps existent seuls. Deux faits peuvent être mentionnés finalement en faveur de l'hypothèse que nous formulons ici :

- jamais dans « La prévision » de Finetti n'outrepasse l'ontologie très restrictive de Hume ;
- Hume constitue une référence commune des tenants d'une interprétation subjectiviste du calcul des probabilités.[79]

Si le subjectivisme en philosophie des probabilités n'engage pas, à la rigueur, à une ontologie humienne, c'est bien dans cette direction que l'ensemble des positions ontologiques associées au subjectivisme tend à s'étendre.

Relativement aux questions d'ontologie, la comparaison avec les théories fréquentiste et logique des probabilités fait apparaître l'originalité propre du propensionnisme : il suppose l'acceptation d'entités naturelles non réductibles aux points d'espace-temps. Mais l'ontologie propensionniste se démarque au premier chef de l'ontologie subjectiviste. Résumons nos arguments à l'appui de cette thèse :

- l'interprétation subjectiviste rend compte du concept de probabilité sans recourir à des entités ontologiquement irréductibles aux points d'espace-temps ;
- en-deçà de la définition opérationaliste des probabilités dans les situations de pari, le refus de faire appel à de telles entités est au fondement de la position subjectiviste en philosophie des probabilités ;
- ce refus s'étend naturellement à une ontologie générale semblable à celle de Hume – qui constitue le type même des ontologies minimalistes.

[77] De Finetti (1937) p. 65.

[78] Voir Hume (1748) section II : « Origine des idées ».

[79] Mentionnons en particulier le concept de « survenance humienne » introduit par Lewis.

Pour finir, remarquons que la prise en compte de l'ontologie subjectiviste permet de requalifier l'ontologie propensionniste, relativement à une question indépendante de celle de l'acceptation d'entités irréductibles aux points d'espace-temps. Pour le subjectiviste en philosophie des probabilités, en effet, la probabilité n'est rien d'autre qu'un degré de croyance. Cette position conduit à soutenir que seuls existent les individus capables d'avoir des idées. Le réalisme ontologique[80] – c'est-à-dire la thèse selon laquelle il existe une réalité objective indépendante des individus susceptibles de percevoir et de penser – contribue donc à caractériser l'ontologie propensionniste. Cette caractérisation et l'opposition des théories propensionniste et subjectiviste qui lui est associée permettent de fonder une analyse de la dimension épistémologique du propensionnisme. Nous y venons maintenant.

5.2.2 Épistémologie propensionniste

Nous venons d'annoncer notre intention d'analyser les présupposés épistémologiques du propensionnisme à partir de la requalification du débat ontologique entre propensionnistes et subjectivistes dans les termes du réalisme et de l'anti-réalisme. Il nous revient alors de montrer comment ce débat ontologique se prolonge au plan de l'épistémologie.

5.2.2.1 Réalisme propensionniste et anti-réalisme subjectiviste. De l'ontologie à l'épistémologie

Les corrélats épistémologiques du réalisme ontologique sont immédiats. Affirmer que le monde est ontologiquement indépendant des individus capables de pensée et de perception ouvre en effet la possibilité d'une divergence entre d'une part le monde tel que nos meilleures théories le décrivent, et d'autre part le monde tel qu'il est réellement. Le projet de connaissance peut alors s'entendre en un sens fort : celui de dire le monde tel qu'il est indépendamment des hommes qui pensent et perçoivent. Pour l'anti-réaliste, à l'inverse, le monde n'existe pas indépendamment des individus capables de pensée et de perception. Le monde ne peut donc pas différer de ce que nous en pensons et percevons ; il est parfaitement décrit par nos meilleures théories, ou plutôt : il est ce que nos meilleures théories décrivent. Dans l'introduction à *The Logical Basis of Metaphysics*, Dummett suggère que la question de la possibilité d'une divergence entre les choses telles qu'elles sont et ce que nous en disons – c'est-à-dire entre le vrai et le tenu pour vrai –

[80] Notre terminologie diffère ici sensiblement de celle de Popper. Dans Popper (1983) il parle en effet de « réalisme métaphysique » au sens où nous parlons ici de « réalisme ontologique ».

est la question du réalisme épistémologique telle qu'elle se formule aujourd'hui. Le débat ontologique entre propensionnistes et subjectivistes à propos du réalisme doit donc se prolonger au plan de l'épistémologie. Qu'en est-il effectivement – c'est-à-dire si l'on s'intéresse directement aux épistémologies propensionniste et subjectiviste ? Il convient de l'étudier maintenant.

La position subjectiviste défendue dans de Finetti (1937) peut être considérée comme plus radicale encore que la position anti-réaliste que nous décrivions dans le dernier paragraphe. En effet, chez de Finetti, il ne s'agit plus seulement d'impossibilité que le vrai diffère du tenu pour vrai. Il s'agit plus radicalement de ceci que la question de la vérité des énoncés probabilistes n'est pas pertinente. Aussi, un énoncé de probabilité ne peut en aucun cas prétendre à la vérité. Il reste « une opinion, qui est et ne peut être autre chose qu'une opinion, donc ni vraie, ni fausse »[81].

Dans de Finetti (1937), l'affirmation selon laquelle les énoncés probabilistes ne sont susceptibles ni de vérité, ni de fausseté est clairement pensée comme une conséquence de l'impossibilité de montrer empiriquement qu'un tel énoncé est erroné :

> un événement quelconque ne peut qu'arriver ou ne pas arriver, et ni dans un cas ni dans l'autre on ne peut décider quel était le degré de doute avec lequel il était "raisonnable" ou "juste" de l'attendre avant de savoir s'il était réalisé ou non.[82]

En particulier, les fréquences relatives observables ne sont pas propres à falsifier les hypothèses probabilistes :

> aucune relation entre probabilités et fréquences n'a de caractère empirique, car la fréquence observée, quelle qu'elle soit, est toujours compatible avec toutes les opinions concernant les probabilités respectives.[83]

Cette idée ne prend sens qu'à la lumière de la loi des grands nombres. Nous avons vu qu'elle a pour conséquence que toute hypothèse relative à la valeur d'une probabilité objective, même vraie, a une probabilité finie non nulle d'être falsifiée par un test fréquentiel. C'est cette non-falsifiabilité des énoncés de probabilité qui conduit de Finetti à considérer que ces énoncés ne peuvent être ni faux, ni – par conséquent – vrais. Les conséquences qu'il tire de la loi des grands nombres conduisent donc le subjectiviste à un anti-réalisme épistémologique radical – au sens où la question d'une divergence entre le vrai et le tenu pour vrai ne peut même pas être posée.

[81] De Finetti (1937) p. 63. Une large part du propos de de Finetti consiste alors à rendre compte de l'idée selon laquelle les observations et expériences passées permettraient de « corriger » les évaluations de probabilités.

[82] De Finetti (1937) p. 18.

[83] De Finetti (1937) pp. 23–24.

De façon similaire, la position popperienne relativement à la loi des grands nombres ne va pas sans la thèse du réalisme épistémologique. Comme nous l'avons indiqué dans le paragraphe 5.1.1.2, l'introduction de critères de falsifiabilité méthodologique autorise à considérer que certaines observations de fréquences relatives falsifient certaines hypothèses relatives aux probabilités objectives. Les observations en question établissent que des énoncés tenus pour vrais sont, en fait, faux. La réponse de Popper à la loi des grands nombres porte donc la marque de l'engagement du propensionnisme au réalisme épistémologique.

Notons pour finir que le désaccord entre subjectivistes et propensionnistes relativement aux conséquences à tirer de la loi des grands nombres suppose une commune prise en compte du problème épistémologique qu'elle pose. Sur ce point, subjectivisme et propensionnisme s'opposent ensemble au fréquentisme qui, définissant les probabilités comme des limites de fréquences, renonce à rendre compte du caractère probabiliste de la convergence des fréquences relatives vers les probabilités.

L'opposition des propensionnistes et des subjectivistes sur la question du réalisme ontologique se prolonge donc bien au plan épistémologique. Le propensionnisme suppose l'adhésion à la thèse épistémologique réaliste de la possibilité d'une divergence entre le monde tel qu'il est réellement et le monde tel que le décrivent nos théories même les meilleures. Le subjectivisme, quant à lui, est tenu à une forme d'anti-réalisme épistémologique, mais plus extrême que celle que le seul débat ontologique nous conduisait à envisager. En-deçà de la question de la divergence entre le vrai et le tenu pour vrai, le fait même que les énoncés relatifs aux probabilités puissent prétendre à la vérité est nié. Cette opposition épistémologique procède d'une divergence relative aux conséquences à tirer de la loi des grands nombres.

La question du réalisme ou de l'anti-réalisme est fondamentale au plan de l'épistémologie ; la plupart des grandes questions épistémologiques lui est subordonnée, ou au moins associée. Nous nous proposons donc d'étudier et de comparer les conséquences épistémologiques du réalisme propensionniste d'une part et de l'anti-réalisme subjectiviste d'autre part. Ce faisant, nous énoncerons quelques thèses caractéristiques de l'épistémologie propensionniste.

5.2.2.2 Épistémologies propensionniste et subjectiviste. Trois points de comparaison

Nous avons vu l'anti-réalisme ontologique conduire de Finetti à considérer que les énoncés probabilistes sont des opinions, et se situent donc hors du

champ de la vérité et de la fausseté. Est-ce à dire que toutes les évaluations de probabilités sont également admissibles ? Il est clair que non : seules sont acceptables les évaluations « cohérentes », c'est-à-dire contre lesquelles il est impossible d'engager un pari « en s'assurant de gagner à coup sûr »[84]. La question de la vérité cède alors la place à celle de la cohérence – logique et pratique. Le prédicat "vrai" est absent de l'épistémologie de de Finetti, au moins en ce qui concerne les probabilités.

On pourrait facilement réintroduire ce prédicat central, en souscrivant à une conception de la vérité comme cohérence. Cette solution est évidemment *ad hoc* ; elle ne résout rien puisque le concept de vérité reste réductible à celui de cohérence. Néanmoins, elle indique le fait épistémologique ici central : le subjectiviste ne peut pas réintroduire la vérité conçue traditionnellement comme correspondance entre le discours et la chose. Comment, en effet, envisager une telle correspondance si l'existence même de la chose – c'est-à-dire de la chose en tant qu'elle est indépendante de celui qui la pense – est niée ? L'anti-réalisme ontologique, celui-là même qui conduit à nier la possibilité d'une divergence entre le vrai et le tenu pour vrai, place le subjectiviste face à l'alternative suivante : soit s'en tenir à la conception traditionnelle de la vérité et souscrire à une épistémologie qui fait l'économie du prédicat "vrai" – ce que fait de Finetti –, soit redéfinir le concept de vérité, la conception qui semble s'imposer alors étant celle de vérité-cohérence.

L'attachement à la conception traditionnelle de la vérité apparaît alors comme un véritable engagement épistémologique. Or, le réalisme épistémologique va de pair avec cet engagement : accepter la possibilité d'une divergence entre le vrai et le tenu pour vrai, c'est supposer que la notion de vérité ne peut pas être définie seulement en termes de discours (ou de croyance). La conception traditionnelle s'impose alors. Ainsi Popper écrit-il dans les toutes premières pages de Popper (1990) :

> c'est de Tarski que j'appris la force de l'idée de vérité absolue et objective, et le fait qu'elle était logiquement défendable ....
> Il s'agit d'une théorie de la vérité *objective* autrement dit de la vérité entendue comme la "correspondance" d'un énoncé avec les faits, et *absolue* : si un énoncé formulé de manière non ambiguë est vrai dans un langage, toute traduction correcte de cet énoncé dans un autre langage est également vraie.[85]

L'adhésion à la thèse de l'objectivité de la vérité (c'est bien elle, et non celle de l'absoluité, qui nous intéresse ici) caractérise donc l'épistémologie propensionniste. Elle est supposée par le réalisme épistémologique..

[84] De Finetti (1937) p. 7.
[85] Popper (1990) p. 22.

Incapable de penser la vérité et la fausseté des évaluations de probabilités, le subjectiviste radical admet exactement les évaluations de probabilités cohérentes. Elles seules, en effet, prémunissent contre la possibilité d'un pari qui sera perdu quoi qu'il arrive, un pari hollandais (un *dutch book*). Dans cette mesure, elles sont conformes aux exigences minimales de la raison. Nous parlerons en ce sens de la cohérence comme critère subjectiviste de la rationalité minimale des évaluations de probabilités. Maintenant, ce critère est *structurel* : il ne concerne pas les évaluations de probabilités individuelles – c'est-à-dire considérées isolément les unes des autres –, mais des ensembles d'évaluations de probabilités. Plus précisément, il apparaît dans de Finetti (1937) que la question de la cohérence se pose pour les ensembles d'évaluations des probabilités d'événements incompatibles dont exactement un doit se réaliser. Un tel ensemble est cohérent si et seulement si la somme des probabilités évaluées est égale à 1 – c'est-à-dire s'il n'invalide pas le théorème des probabilités totales.

De façon plus générale, le théorème « de Ramsey – de Finetti » établit que la cohérence équivaut, pour un ensemble d'évaluations de probabilités, à la satisfaction des axiomes de Kolmogorov pour le calcul des probabilités.[86] Ce théorème confirme que les contraintes de rationalité véhiculées par le subjectivisme radical sont minimales : en un sens, on ne peut pas exiger moins d'une théorie qui prétend interpréter le calcul des probabilités que la satisfaction des axiomes de ce calcul. Il prouve aussi que ces contraintes sont exclusivement structurelles : si l'on fait abstraction de l'axiome conventionnel selon lequel les probabilités prennent leur valeur dans l'intervalle $[0\,;\,1]$, aucun des axiomes du calcul des probabilités ne porte sur les évaluations de probabilités particulières. En faisant de la cohérence le critère de la rationalité, le subjectiviste ne peut donc poser la question de la rationalité des évaluations de probabilités qu'au plan structurel.

A l'inverse, la question de la rationalité d'une évaluation de probabilité prise *individuellement* a un sens pour le propensionniste. La thèse d'une divergence possible entre le vrai et le tenu pour vrai le conduit à concevoir la rationalité comme le fait de mettre tout en oeuvre pour minimiser cette divergence. Dans le cas particulier de l'évaluation de la probabilité d'un événement E, la rationalité revient à prendre en compte tous les éléments d'information pertinente disponibles relativement à la classe d'événements définie par les conditions d'engendrement de E. Les fréquences observées jouent alors le premier rôle, sinon le seul. La rationalité est donc d'abord une question locale – c'est-à-dire qui se pose pour chaque évaluation de probabilité prise individuellement.

[86] L'axiome d'additivité dénombrable n'est cependant pas pris en compte par ce résultat.

Le critère subjectiviste de rationalité des évaluations probabilistes n'est pas exclusif : pour le même événement, plusieurs évaluations de probabilité différentes sont susceptibles d'y satisfaire. C'est même toujours le cas dès lors qu'on considère au moins deux événements incompatibles dont l'un exactement va se réaliser. Le choix d'un ensemble d'évaluations de probabilités parmi ceux qui sont acceptables est alors « arbitraire »[87], irréductiblement subjectif :

> chacune de ces évaluations correspond à une opinion cohérente, à une opinion légitime en soi, et chaque individu est libre d'adopter celle de ces opinions qu'il préfère, ou, pour mieux dire, celle qu'il sent.[88]

Il devient difficile, dans ces conditions, d'expliquer que des individus différents puissent choisir des évaluations de probabilités identiques. Nous avons indiqué déjà que de Finetti (1937) est largement consacré à rendre intelligible le phénomène fréquent de la coïncidence des opinions subjectives, à « rendre compte des concordances plus ou moins strictes qu'on observe entre les jugements des divers individus ainsi qu'entre les prévisions et les résultats observés »[89].

Le statut des « concordances ... qu'on observe entre les jugements des divers individus » est tout autre dans le cadre réaliste de la théorie propensionniste des probabilités. De Finetti reconnaît que ces « concordances » sont immédiatement expliquées si l'on admet « l'existence d'une probabilité objective » : les opinions individuelles coïncident alors car elles se réfèrent à la même entité. Cet argument n'est valide que si l'on suppose que les différents individus se réfèrent à la probabilité objective commune sur un mode commun : le fait qu'ils parlent tous de la même propension – et donc de la même probabilité – n'implique pas qu'ils en disent la même chose. Or, la thèse réaliste selon laquelle le tenu pour vrai peut différer du vrai impose – tout en la rendant intelligible – l'idée selon laquelle différents individus pourraient se référer à la même probabilité sur des modes différents. Si les opinions subjectives relatives à la valeur d'une même probabilité objective coïncident, c'est que la divergence entre cette probabilité et son évaluation est la même chez tous les individus. Le plus raisonnable est alors de supposer que ce rapport est de correspondance, autrement dit que l'évaluation de probabilité partagée est vraie. Dans le cadre de l'épistémologie réaliste du propensionnisme, la coïncidence des opinions subjectives a donc le statut d'un indice de la vérité. Elle est bien moins un phénomène empirique dont il faut parvenir à rendre compte, qu'une situation épistémologiquement recherchée.

[87] De Finetti (1937) p. 8.
[88] De Finetti (1937) p. 8.
[89] De Finetti (1937) p. 16.

En prenant appui sur l'opposition des épistémologies subjectiviste et propensionniste, nous avons mis en évidence trois thèses caractéristiques de l'épistémologie propensionniste et subordonnées à celle du réalisme :

- la vérité se définit comme correspondance entre la réalité et ce qu'on en dit ;
- la rationalité est celle des énoncés probabilistes considérés individuellement plutôt que celle de la structure qu'ils composent ensemble ;
- la coïncidence des opinions subjectives est plus qu'un phénomène empirique qu'il faut expliquer : elle est l'indice recherché de la probable vérité de l'énoncé sur lequel l'accord se fait.

Ce faisant, nous achevons notre analyse des dimensions ontologique et épistémologique du propensionnisme. Il est apparu que, dans ces deux champs, le propensionnisme suppose des prises de position dans des débats structurants – notamment celui du réalisme et de l'anti-réalisme – et qu'il se distingue par là des autres théories des probabilités en général et du subjectivisme en particulier. En vue d'affiner ce verdict, nous nous tournons finalement vers la métaphysique.

5.2.3 Métaphysique propensionniste

Nous appelons ici « métaphysique » toute théorie des principes de l'être – avec les mots de Popper : un ensemble « de conceptions générales de la structure du monde et, en même temps, de conceptions générales de la situation de problème dans la cosmologie physique »[90]. Comprises comme entités méta-physiques, inobservables, les propensions sont caractérisées par leur activité en vue de la réalisation de tel ou tel événement possible. Dans cette mesure, elles s'inscrivent dans une tradition inaugurée par le concept aristotélicien de puissance. Plus précisément, les propensions peuvent être rapprochées des « puissances irrationnelles » qui – contrairement aux puissances « rationnelles » dont elles sont distinguées au livre Θ de la *Métaphysique* – se manifestent en effet « dans les êtres inanimés » et sont « puissances d'un seul effet ». Le concept de puissance fonde chez Aristote une métaphysique au sens où nous avons défini ce terme. En outre, Popper y fait souvent référence, et toujours finalement pour s'en démarquer. Dans ces conditions, comparer les concepts de puissance et de propension doit être fécond à ce point de notre étude.

[90] Popper (1982a) p. 161.

5.2.3.1 Puissances et propensions

Nous venons d'indiquer qu'il est assez fréquent que Popper, cherchant à caractériser les propensions, en arrive à évoquer la notion aristotélicienne de puissance. La remarque qu'il fait alors est toujours la même :

> Comme toute propriété dispositionnelle, les propensions présentent une certaine ressemblance avec les potentialités aristotéliciennes. Mais il y a une différence de taille : contrairement à ce que les aristotéliciens seraient peut-être enclins à croire, les propensions ne sauraient être inhérentes aux *choses* individuelles.[91]

Noter cette différence est un moyen pour Popper de mettre en avant l'idée fondamentale de sa théorie des probabilités, qui consiste à les penser comme des propriétés d'ensembles de conditions physiques. Mais limiter à cette remarque le rapprochement des propensions et des puissances aristotéliciennes, comme le fait Popper, est insuffisant. Cela laisse entendre, sans pourtant le justifier, d'une part qu'il n'y a pas d'autres différences significatives entre les deux théories, d'autre part que la différence mise en évidence invalide l'idée même de leur rapprochement. Or, nous verrons en poussant plus avant la comparaison des puissances et des propensions que cette suggestion est deux fois trompeuse.

Avant de montrer en quoi le rapprochement des propensions et des puissances est bien fondé et permet de penser certains aspects de la métaphysique à laquelle le propensionnisme engage, il convient d'en finir avec ce qui sépare les deux théories. Nous avons vu que la différence notée par Popper est importante en tant qu'elle porte sur un point central du propensionnisme. Elle n'est toutefois pas seule dans ce cas ; trois autres différences de cet ordre doivent être notées :

- Aristote ne pense jamais la puissance comme une entité séparée. Non seulement la puissance n'existe pas indépendamment de la chose dont elle est puissance, mais encore – et surtout – elle n'existe que comme qualification de l'être de cette chose. La puissance est un mode de l'être ; l'utilisation propre du terme « puissance » est celle qui en faite dans l'expression « être en puissance » ;
- contrairement aux propensions – dont le degré de réalité est mesuré par les probabilités – , le degré de puissance des êtres aristotéliciens n'est pas susceptible de quantification. Il est d'ailleurs évident que la question ne se pose même pas à Aristote. Cela nous amène à la troisième des différences que nous relevons :
- la théorie aristotélicienne du changement et le propensionnisme n'ont

[91] Popper (1983) p. 372.

pas le même statut. La notion de puissance entre chez Aristote dans une théorie du changement. Plus précisément, le couple de l'acte et de la puissance est au principe de tous les changements que permettent de décrire les concepts de forme, de matière et de privation de forme – et qu'*explique* la théorie des causes. La notion de puissance apparaît donc dans la partie *descriptive* d'une théorie du changement. A l'inverse, le propensionnisme n'est pas d'abord une métaphysique, mais une théorie des probabilités. Il se veut en outre une théorie explicative, censée rendre compte de la stabilité et de la tendance à la convergence des fréquences observées dans une suite de répétitions indépendantes d'une expérience aléatoire. Quoi qu'il en soit plus précisément de cette question, le propensionnisme n'est pas conçu par Popper comme une théorie descriptive appelant le complément d'une théorie explicative. Ainsi la question de l'actualisation des propensions – c'est-à-dire de la réalisation des événements qu'elles tendent à faire advenir – n'est-elle pas du tout centrale. Elle est abordée une seule fois à notre connaissance, et brièvement :

> Je crois quant à moi que le moteur de ce processus est une combinaison d'*accidents* et de *préférences* : les préférences des organismes, dans leur quête d'un monde meilleur, pour certaines possibilités.[92]

Il nous semble impossible de considérer cette phrase comme une théorie de l'actualisation des propensions.

Finalement, il était bien insuffisant de limiter l'analyse du rapprochement entre propensions et « potentialités aristotéliciennes » à l'idée selon laquelle les unes sont attachées à des ensembles de conditions physiques et les autres à des choses. Nous venons en effet de mettre en évidence trois autres différences, qui toutes trois touchent aussi à des aspects fondamentaux du propensionnisme. Ces différences notées, peut-on encore penser sérieusement le rapprochement – qui nous était pourtant apparu évident – du propensionnisme considéré dans ses aspects métaphysiques et de la théorie aristotélicienne du changement ? A cette question, nous répondons par l'affirmative. Nous considérons que les différences qui séparent ces deux théories ne doivent pas masquer – et rendent même d'autant plus remarquable – la communauté de leur inspiration. Plus précisément, toutes deux pourraient être décrites comme des « métaphysiques du changement »[93]. Dans le prochain paragraphe, nous nous proposons de donner sens à cette expression.

[92] Popper (1990) pp. 49–50.

[93] L'expression est de Renée Bouveresse. Elle l'utilise dans Bouveresse (1981) (p. 128) pour qualifier la métaphysique de Popper.

5.2.3.2 Une métaphysique du changement

Revenons, pour commencer, sur la notion de possibilité. Celle-ci est centrale dans chacune des deux théories qui nous occupent. D'un côté, un être est en puissance tout ce qu'il *peut* être ; de l'autre, une propension tend activement à l'actualisation d'un événement *possible*. En outre, elle est considérée par chacune dans sa relation – et non sa simple opposition – à la notion de réalité. Enfin, surtout, dans les deux cas, cette relation conduit d'abord à attribuer une forme de réalité au possible. Montrons-le rapidement.

Pour ce qui est d'Aristote, il annonce[94] deux solutions au problème du changement tel qu'il se pose après la critique éléatique. Elles reposent l'une sur la distinction par soi / par accident, l'autre sur la distinction acte / puissance. C'est la seconde qui nous intéresse ici. Elle consiste à penser que l'être est déjà en un sens – celui de la puissance – ce qu'il va devenir. Or, « une chose est possible si, quand elle passe à l'acte dont elle est dite avoir la puissance, il n'en résulte aucune impossibilité »[95]. Le critère du possible dans l'être est la possibilité *logique* de l'actualisation de la puissance. Autrement dit, le possible dans une chose est ce qui mérite d'être dit en puissance. Du coup, il a bien une forme sinon d'actualité, du moins de réalité : « le possible est ... dans la mesure où son actualisation est possible »[96].

Pour ce qui est de Popper, nous l'avons vu décrire les propensions comme des « possibilités ... [qui] ont une réalité physique ». Cette expression semble suffire à montrer que le possible est conçu, en un sens, comme réel. Seulement, il est apparu qu'elle est de ces expressions confuses qui obscurcissent la notion de propension chez Popper. A l'ensemble de ces expressions, nous avons proposé de substituer une description plus claire et qui en un sens ferait droit à chacune. A la lumière de cette description, l'idée selon laquelle les propensions sont des « possibilités ... [qui] ont une réalité physique » doit être comprise ainsi : les propensions sont réelles en tant que 1. elles sont actives et 2. elles peuvent actualiser des possibilités. Il apparaît alors que Popper pense une forme de réalité des événements possibles[97], qui réside tout entière dans les propensions agissant en vue de leur réalisation.

Ainsi le propensionnisme et la théorie aristotélicienne du changement conduisent-ils tous deux à attribuer une réalité au possible. Cette

[94] Aristote (*Physique*) I, 8.

[95] Aristote (*Métaphysique*) Θ, 3, 1047a 25, pp. 31.

[96] Aristote (*Métaphysique*) Θ, 4, 1047b 3, p. 32.

[97] On ne confondra pas cette réalité avec la réalité modale que certains philosophes accordent aux mondes possibles. Le réalisme modal des mondes possibles et le réalisme de Popper ne sont ni des thèses équivalentes, ni même des thèses solidaires l'une de l'autre. Ainsi, Popper n'est pas un réaliste modal : il n'admet qu'une réalité. A l'inverse, le réalisme modal de Lewis s'accompagne d'une conception subjectiviste des probabilités.

thèse commune ne suffit pas en elle-même à établir l'unité d'inspiration des métaphysiques aristotélicienne et popperienne. Pour achever la démonstration, il faut montrer comment cette thèse prend place dans une théorie de l'être commune à Aristote et à Popper. Plus explicitement, chacun des deux auteurs pense l'existence comme la réalité du possible portée à un degré extrême.[98] L'être réel est alors le possible en tant qu'il a acquis ce degré de réalité, qu'il est devenu actuel. En ce point précis, l'analogie entre la puissance et les propensions peut être réintroduite par-delà les différences importantes qui séparent les deux notions. En effet, l'une et les autres jouent – respectivement dans la théorie aristotélicienne et dans la théorie popperienne – des rôles analogues : elles sont le principe actif qui porte le possible à la réalité. Tout ce qui est réellement, est du possible actualisé, par la puissance chez Aristote, par les propensions chez Popper. Il apparaît alors que le propensionnisme popperien est, de même que la théorie aristotélicienne du changement, une métaphysique. Nous reprenons l'expression « métaphysique du changement » pour les penser ensemble comme métaphysiques qui 1. accordent une forme de réalité au possible ; 2. pensent l'existence comme cette réalité portée à son degré extrême.

La grande similitude des conceptions aristotélicienne et popperienne du réel redonne sens au projet d'analyse de la métaphysique propensionniste à la lumière du rapprochement de la puissance et des propensions. Plus précisément, nous nous appuierons sur la *Physique* d'Aristote pour montrer comment certaines thèses métaphysiques relatives à la nature – Popper dirait « la structure du monde » – sont engagées par une métaphysique du changement. Ce texte est en effet ici un bien meilleur guide que les passages de l'oeuvre de Popper dans lesquels des considérations métaphysiques sont livrées sans que soient clarifiés les liens conceptuels qui les unissent. La métaphysique du changement permet à Aristote de penser le changement comme un « devenir », c'est-à-dire comme :

– ni chaotique : la métaphysique du changement est pour Aristote un moyen de

> comprendre que parmi tous les étants aucun, par nature, ni ne fait n'importe quoi, ni ne subit [n'importe quoi] du fait de n'importe quoi, pas plus que n'importe lequel vient de n'importe quel autre, à moins qu'on ne l'entende par accident.[99]

En effet, si tout ce qui est, est en vertu de l'actualisation d'une possibilité et si, d'autre part, une puissance est puissance d'un seul effet,

[98] L'idée de quantifier le degré de réalité du possible, et en particulier de le quantifier sur l'intervalle [0 ; 1], paraît alors assez naturelle.

[99] Aristote (*Physique*) I, 5, 188a 34, p. 93.

un individu ne peut pas devenir n'importe quoi. Plus généralement, n'importe quoi ne peut pas advenir ; le changement n'a lieu qu'entre des bornes fixées par la configuration initiale. Le monde lui-même n'est donc pas chaotique, mais profondément structuré ;

– ni pour autant complètement déterminé : il ne « résulte aucune impossibilité »[100] non seulement de l'actualisation du possible, mais encore de sa non-actualisation : « rien n'empêche qu'une chose capable d'exister ou de devenir, en fait ne soit, ni ne se réalise »[101]. A strictement parler, on peut conclure d'une métaphysique du changement à l'indétermination *logique* du changement : jusqu'au moment où il a lieu, il est logiquement possible qu'il n'ait pas lieu. Mais notons bien que l'indétermination dont la notion de possible est porteuse chez Aristote n'est pas seulement logique. Elle est inscrite dans le possible lui-même en tant qu'être, et plus précisément dans sa matière. Elle ne pourra être réduite par aucun progrès de la science, aussi grand soit-il : la matière est par principe inintelligible. La nature est donc essentiellement indéterministe.

Nous avons montré que la métaphysique dont le propensionnisme est corrélatif et la théorie aristotélicienne du changement dans l'être peuvent être pensés comme deux métaphysiques d'un même type. Nous proposons de parler de métaphysique du changement. Il est apparu qu'une telle métaphysique conduit à concevoir le changement comme un devenir au sens aristotélicien, et donc à penser le monde comme à la fois profondément structuré et indéterministe. En précisant en ce sens la notion de métaphysique du changement, nous ne prétendons pas avoir découvert des présupposés inconscients, tus, ou non assumés, de la pensée de Popper. Nous avons au contraire rejoint des thèmes métaphysiques chers à cet auteur et amplement développés dans la partie de son oeuvre consacrée au propensionnisme. Notre ambition est donc plutôt d'avoir contribué à clarifier les corrélats métaphysiques du propensionnisme, en analysant les liens conceptuels qui unissent certaines des thèses popperiennes en ce domaine. Ce projet mérite maintenant d'être poursuivi dans le sens d'une analyse de ce qui fait l'originalité de la métaphysique corrélative de la théorie propensionniste des probabilités (« métaphysique propensionniste » dans la suite).

[100] Aristote (*Métaphysique*) Θ, 3, 1047a 25, p. 31. Nous avons déjà cité ce passage.
[101] Aristote (*Métaphysique*) Θ, 4, 1047b 9, p. 32.

5.2.3.3 Originalité de la métaphysique propensionniste

Nous relevons trois points par où la métaphysique propensionniste de Popper se distingue significativement de la théorie aristotélicienne du changement dans l'être. Nous nous proposons de présenter successivement ces différences locales et de montrer, pour chacune, qu'elle peut être pensée en termes de modernité de la version popperienne de la métaphysique du changement.

Rapport entre les deux aspects du devenir. Alors qu'Aristote ne le fait pas, Popper pense le rapport entre les deux caractéristiques du monde tel que la métaphysique du changement conduit à le concevoir. Pour le dire plus précisément, au-delà de la simple juxtaposition des deux concepts d'indéterminisme et de déterminisme relatif auxquels la métaphysique du changement conduit, Popper pense l'articulation de ces deux concepts.

Dans la dernière partie de Popper (1982a), Popper développe en effet l'idée selon laquelle le déterminisme scientifique vaudrait à titre d'approximation de l'indéterminisme réel du monde. Son argumentation s'appuie alors sur le caractère déterministe de la relation qui unit les propensions présentes aux situations passées. Le surcroît d'analyse que nous pointions chez Popper dépend donc tout entier de la notion de déterminisme scientifique. Or, cette notion est récente. On la fait le plus souvent remonter à l'*Essai philosophique sur les probabilités* de Laplace publié en 1814 ; il est en tout cas clair qu'elle n'est pas antérieure à l'âge classique.

Métaphysique du changement et indéterminisme métaphysique. Il est un autre point sur lequel l'analyse de Popper est plus poussée que celle d'Aristote : celui de l'articulation de la thèse de l'indéterminisme métaphysique à la métaphysique du changement. Nous avons montré que la théorie de l'être comme possible actualisé implique chez Aristote un indéterminisme *logique* – le possible est caractérisé logiquement par ceci qu'il peut être et qu'il peut ne pas être. Puis nous avons vu qu'Aristote, sans s'en justifier et sans paraître même s'en inquiéter, va au-delà de cette thèse et pense grâce à la notion de matière l'indéterminisme *métaphysique* de la nature.

Sur ce point précis, Popper (1990) défend une position comparable à celle d'Aristote. Popper aussi dépasse le point de vue de l'indéterminisme logique – ou, pour employer une terminologie plus contemporaine, scientifique. Ainsi écrit-il : « les propensions en devenir sont des processus objectifs qui n'ont

rien à voir avec notre manque d'information »[102]. Surtout, Popper thématise dans ce texte le lien qui unit la thèse propensionniste à la thèse métaphysique indéterministe. Plus précisément, il le fait apparaître comme comparable à celui qui unit un moyen à une fin qu'il soutient.

Il n'en a pas toujours été ainsi dans l'oeuvre de Popper. En particulier, l'abandon du déterminisme dit « scientifique » est présenté dans Popper (1982b) comme une condition de possibilité de l'hypothèse propensionniste :

> C'est seulement en écartant le déterminisme que nous obtiendrons la liberté nécessaire pour considérer sérieusement l'interprétation de la théorie des propensions en tant que théorie physique.[103]

Nous retenons ici l'analyse proposée dans Popper (1990) car nous considérons que c'est dans ce seul texte que Popper déploie sciemment la dimension métaphysique qui sous-tend le propensionnisme.

L'idée alors développée, selon laquelle l'indéterminisme serait une fin, ne trouve une justification chez Popper que dans la thèse récurrente de l'indéterminisme comme condition de possibilité de la créativité pratique et – surtout – de la liberté métaphysique. Or, le concept de liberté métaphysique est un concept éminemment moderne, absent de la philosophie antique en particulier.

Par deux fois, nous avons pointé un surcroît d'analyse du propensionnisme par rapport à ce que l'aristotélisme classique propose. Dans le dernier point que nous nous proposons d'analyser, les choses diffèrent sensiblement : le propensionnisme propose une réponse différente de celle de l'aristotélisme à une question laissée ouverte dans le cadre partagé de la métaphysique du changement. Cette question est celle de la possibilité de la nouveauté.

Possibilité de la nouveauté. Pour comprendre la divergence des réponses aristotélicienne et popperienne à la question de la possibilité de la nouveauté, il convient de revenir à la seule différence entre les propensions et les potentialités aristotéliciennes qui est discutée par Popper. Les propensions sont des propriétés d'ensembles de conditions physiques tandis que la puissance est une propriété individuelle des êtres animés ou inanimés. Une conséquence du caractère individuel de la puissance est qu'elle suppose pour s'exercer un être singulier dans lequel aura lieu le changement en vue duquel elle agit. Dans

[102]Popper (1990) p. 40. Par une telle affirmation, Popper entend distiguer la thèse qu'il défend des interprétations épistémiques des probabilités – qui associent la notion de probabilité à l'idée d'un défaut de connaissance – aussi bien que de l'hypothèse métaphysique déterministe.

[103]Popper (1982b) p. 78.

ces conditions, il apparaît que la métaphysique aristotélicienne ne peut pas accorder de vrai statut à la nouveauté. Celle-ci n'est pensée que de manière marginale – sans doute sous les deux catégories de monstre et de contingence – mais reste, justement, monstrueuse. Elle n'est pas susceptible d'être intégrée au régime normal de fonctionnement de la nature.

Il en va tout autrement avec le propensionnisme. Les possibilités dont les propensions tendent à augmenter le degré de réalité sont attachées à des ensembles de conditions physiques. Le changement ne suppose donc pas un individu qu'il affecterait. Il semble donc possible que l'apparition de nouveaux types d'êtres soit de ces possibilités que certaines situations physiques peuvent déterminer, et l'activité des propensions porter à l'actualité.

Mais Popper va plus loin, posant une sorte de principe de plénitude :

> toutes les possibilités non nulles, y compris celles auxquelles sont attachées des propensions très petites, finiront par s'actualiser pourvu qu'elles aient le temps de le faire ; autrement dit, pourvu que les conditions pertinentes se répètent assez souvent, ou demeurent constantes sur une durée suffisamment longue.[104]

Avec ce principe, Popper non seulement prend acte de la possibilité de penser la nouveauté dans le cadre propensionniste, mais encore la rend nécessaire. Dès lors que les possibilités nouvellement actualisées déterminent des ensembles de possibilités dont certaines sont vraisemblablement nouvelles, et que ces nouvelles possibilités finiront – selon le principe de plénitude – par être actualisées, Popper s'autorise à écrire que « l'univers de propensions qui est le nôtre est intrinsèquement créatif »[105]. Cet aspect de la métaphysique propensionniste nous paraît constituer sa plus grande originalité par rapport à la pensée aristotélicienne, dont elle semblait par ailleurs très proche. Or, elle peut être réinscrite dans le cadre plus général d'une différence fondamentale entre la pensée antique et la pensée moderne, entre une pensée qui reste – même chez Aristote – une pensée en univers clos et une pensée que la science a ouverte à l'idée de nouveauté. De même que sur les deux autres points que nous avons analysés, l'originalité du propensionnisme suppose des concepts modernes. Dans ces conditions, le propensionnisme peut être pensé comme une métaphysique *contemporaine* du changement.

[104]Popper (1990) p. 42.
[105]Popper (1990) p. 42.

5.3 Conclusion

L'analyse de la dimension philosophique du propensionnisme a montré que cette théorie a d'importants présupposés dans les deux domaines de l'ontologie et de l'épistémologie, par lesquels elle se distingue des autres théories des probabilités. Elle a surtout établi que la dimension philosophique du propensionnisme ne peut pas être pleinement comprise à la lumière de la seule notion de présupposition. Plus qu'il ne *suppose* une métaphysique, le propensionnisme peut être considéré *comme* une métaphysique, version contemporaine d'une métaphysique de type aristotélicien. Là se trouve indéniablement son originalité philosophique la plus grande, parce que radicale, dans le champ des théories des probabilités.

Ces résultats sont ceux de la section 5.2. Ils ont été établis à la lumière de la présentation du propensionnisme à laquelle nous nous sommes livrés dans la section 5.1. Or, cette présentation est fondée sur une analyse du propensionnisme popperien. La question se pose alors de savoir si certains – et, si oui, lesquels – des résultats de la section 5.2 valent spécifiquement du propensionnisme popperien, à l'exception d'autres propensionnismes. Pour répondre à cette question, il est nécessaire de déterminer ce qui est spécifique du propensionnisme popperien dans la caractérisation de la sous-section 5.1.1.

Dans la caractérisation de la sous-section 5.1.1, une affirmation nous paraît spécifique des propensionnismes de type popperien et en tout cas non partagée par tous les tenants d'une interprétation propensionniste des probabilités : l'affirmation selon laquelle les propensions sont des entités autonomes qui existent physiquement. Contre cette affirmation, on peut adopter une conception non réaliste de ces dispositions que sont les propensions. Cette position est peu fidèle à la lettre des écrits popperiens, mais est plus attrayante, d'un point de vue empiriste, que la position popperienne.

La possibilité de refuser l'existence physique des propensions implique, pour qui s'en saisit, la nécessité de renoncer aux conclusions de la sous-section 5.2.1 consacrée à l'ontologie du propensionnisme popperien. Mais elle n'a pas de conséquence relativement au prochain chapitre. Autrement dit, les développements du prochain chapitre ne dépendent pas de l'hypothèse ontologique de l'existence des propensions. Qu'on considère ou non que les propensions existent physiquement, on pourra suivre ces développements et adhérer à leurs conclusions.

Chapitre 6

Propensionnisme et causalité

Dans le chapitre 5, nous avons présenté la théorie propensionniste des probabilités. Plus précisément, nous avons proposé une caractérisation minimale du propensionnisme popperien (sous-section 5.1.1), puis défendu contre le propensionnisme de long-terme, le propensionnisme de cas singuliers qui nous intéresse (sous-section 5.1.2). Ensuite (section 5.2), nous avons mis au jour les positions philosophiques corrélatives du propensionnisme de cas singuliers. Armés de ces analyses, nous pouvons nous attaquer à la question qui motive la présente partie de notre travail, celle du rapport entre les probabilités objectives d'événements singuliers et la causalité.

De ces probabilités objectives d'événements singuliers, nous avons émis l'hypothèse à l'issue de la sous-section 5.1.3 qu'elles sont interprétées par le propensionnisme. Plus précisément, nous considérons que le propensionnisme tel que nous l'avons présenté dans le chapitre 5 est une interprétation du calcul des probabilités *absolues*. Or, la première partie de notre travail témoigne tout entière de ce que le rapport avec la causalité concerne les probabilités *conditionnelles*. Surtout, c'est bien en termes de probabilités conditionnelles que se comprend la notion d'augmentation de probabilités qu'on trouve au fondement des théories probabilistes de la causalité. Dans ces conditions, l'enquête que nous menons dans cette seconde partie de notre travail exige que nous nous tournions des probabilités absolues vers les probabilités conditionnelles.

Le mouvement par lequel nous nous tournons vers les probabilités conditionnelles nous conduit à apercevoir immédiatemment une difficulté majeure pour le propensionnisme en tant qu'il porterait sur les probabilités conditionnelles. Selon cette difficulté, connue sous le nom de « paradoxe de Humphreys », le propensionnisme ne pourrait pas interpréter les probabilités conditionnelles. Cette difficulté nous intéresse au premier chef ici pour deux raisons : d'abord la raison que nous venons de dire, à savoir qu'on voit mal

comment penser le rapport entre causalité et probabilités en l'absence d'une notion de probabilités conditionnelles ; ensuite parce que le paradoxe de Humphreys procède du rapport que les propensions entretiennent avec les causes. Dans ces conditions, le chapitre qui commence et porte sur le rapport entre propensionnisme et causalité est largement consacré à analyser le paradoxe de Humphreys et à tenter de résoudre.

Dans le détail, le chapitre s'organise de la manière suivante :

1. nous présentons le paradoxe de Humphreys ;
2. nous analysons le désaccord entre Humphreys et McCurdy, qui est l'un des auteurs qui prétendent résoudre le paradoxe. Cette analyse nous permet de mettre au jour les ressorts du paradoxe de Humphreys ;
3. nous proposons une interprétation propensionniste des probabilités conditionnelles ;
4. nous discutons notre proposition d'interprétation des probabilités conditionnelles, et en particulier la question de savoir si elle résout effectivement le paradoxe de Humphreys ;
5. nous concluons relativement à la question du rapport entre la causalité et les probabilités dans le cadre du propensionnisme.

6.1 Propensions conditionnelles et causalité : le paradoxe de Humphreys

Du paradoxe introduit dans Humphreys (1985)[1], il existe plusieurs versions. Certaines de ces versions sont formelles ; d'autres ne le sont pas. Dans la section qui commence, nous présentons d'abord une version informelle du paradoxe, qui se comprend aisément à la lumière du seul chapitre 5 et à laquelle Popper semble bien souvent prêter le flanc. Dans une deuxième sous-section, nous présentons des versions formelles du paradoxe. Une troisième sous-section est consacrée aux tentatives de résolution du paradoxe. Un type de solutions proposées retient alors particulièrement notre attention.

[1] La difficulté est identifiée par Humphreys et connue des philosophes des sciences dès la fin des années 1970 (cf. Salmon (1979)). L'expression « paradoxe de Humphreys » est d'ailleurs introduite dans Fetzer (1981). Mais Humphreys (1985) est l'article par lequel Humphreys publie son analyse.

6.1.1 Une version informelle du paradoxe

6.1.1.1 Caractère relatif des probabilités sous l'interprétation propensionniste

Nous avons affirmé au début du présent chapitre que le propensionnisme tel que nous l'avons présenté est une interprétation des probabilités *absolues*. Cela, pourtant, ne va pas de soi. En effet, et nous y avons insisté dès le paragraphe 5.1.1.1, les propensions sont toujours *relatives* à un ensemble de conditions physiques. Par conséquent, les probabilités définies comme mesures de propensions sont elles aussi toujours relatives à un ensemble de conditions physiques. Il semble en découler que toutes les probabilités sont conditionnelles sous l'interprétation propensionniste :

> puisque les probabilités dépendent de conditions génératrices, toutes les probabilités sont des probabilités conditionnelles, l'événement conditionnant étant la réalisation des conditions génératrices.[2]

Cette position semble être celle de Popper. Celui-ci est convaincu de longue date que les probabilités conditionnelles – qu'il appelle « relatives » – sont en fait les termes primitifs du calcul des probabilités.[3] Surtout, il semble clair depuis les premiers textes consacrés au propensionnisme que l'interprétation vise les probabilités conditionnelles :

> un énoncé de probabilité relative tel que :
>
> $$(2) \qquad p(a|b) = r$$
>
> signifie : "La probabilité de l'événement a dans la situation b (ou étant donné les conditions b) est égale à r".[4]

Les probabilités absolues ne valent alors qu'à titre d'abréviations de probabilités conditionnelles, qui n'ont lieu d'être que « lorsque nous nous intéressons à une situation qui n'évolue pas (ou dont les changements peuvent être négligés) »[5].

6.1.1.2 Probabilités conditionnelles et causalité

Les probabilités conditionnelles qu'interprètent le propensionnisme selon la lecture développée dans le dernier paragraphe ont une forte coloration

[2]Milne (1986) p. 129. Ce raisonnement n'est pas celui de Milne. Il est seulement *décrit* par lui, dans des termes particulièrement clairs, et tels que nous le citons.

[3]On notera d'ailleurs que Popper a proposé plusieurs axiomatisations pour le calcul des probabilités relatives. Voir en particulier les « Nouveaux appendices » à Popper (1934).

[4]Popper (1990) p. 38.

[5]Popper (1990) p. 38.

causale. En effet, entre le conditionnant et le conditionné d'une telle probabilité, il y a une propension. Or les propensions, en tant qu'elles ont une réalité physique et qu'elles sont susceptibles de réaliser des événements singuliers, ressemblent beaucoup à des causes singulières. Popper considère plus précisément les propensions comme des causes d'un type particulier :

> La propension maximale [de mesure 1] correspond au cas particulier d'une force classique en acte : une cause au moment même où elle agit.[6]

En d'autres termes, la notion de propension serait une généralisation de la notion de cause. Le propensionnisme apparaît alors comme une théorie selon laquelle le rapport entre le conditionnant et le conditionné d'une probabilité conditionnelle est le rapport entre un ensemble de conditions physiques et un événement qu'il est susceptible de causer.

Ici se manifeste sous une première forme la difficulté que le propensionnisme rencontre avec les probabilités conditionnelles. En effet, selon l'analyse que nous venons de proposer, le propensionnisme conduit à interpréter le rapport entre le conditionné et le conditionnant d'une probabilité conditionnelle comme une relation de nature causale. Le propensionnisme apparaît alors comme une théorie qui instaure une asymétrie – celle de la causalité – dans la relation entre le conditionnant et le conditionné d'une probabilité conditionnelle. Or, le calcul des probabilités est symétrique relativement à la conditionalisation, au sens où tout ce qu'il implique relativement à un couple (A, B) d'événements de probabilité non nulle, il l'implique relativement à (B, A). Dans ces conditions, il semble clair que le propensionnisme ne peut pas prétendre être une interprétation du calcul des probabilités.

6.1.1.3 Conditionnants fondamentaux et conditionnants d'événements

A ce point, on comprend mal comment le raisonnement que nous venons de mener s'articule avec les développements de la sous-section 5.1.3. Plus spécifiquement, on comprend mal comment ce raisonnement ne nous a pas conduit alors à rejeter fermement l'idée selon laquelle le propensionnisme pourrait être une interprétation du calcul des probabilités. La situation, toutefois, s'éclaircit à la lumière de la distinction entre « probabilités conditionnelles fondamentales » (*fundamental conditional probabilities*) et « probabilités conditionnelles d'événements » (*event-conditional probabilities*) telle qu'elle est thématisée par Gillies.

[6] Popper (1990) p. 34.

Selon Gillies, avec les probabilités conditionnelles fondamentales, « nous avons la forme P(A | S) où S est un ensemble de conditions répétables »[7]. En d'autres termes, les probabilités conditionnelles fondamentales sont les probabilités que nous évoquions dans les deux derniers paragraphes, celles dont le conditionnant est un ensemble de conditions physiques du type de ceux qui déterminent les fonctions de probabilités selon l'interprétation propensionniste. Maintenant,

> de telles probabilités conditionnelles fondamentales peuvent être opposées aux probabilités conditionnelles de la forme P(A | B), où B n'est pas un ensemble de conditions répétables, mais un événement. ... Appelons ces probabilités des *probabilités conditionnelles d'événements*.[8]

Sous la distinction entre conditionnants fondamentaux et conditionnants d'événements, il apparaît que si le propensionnisme conduit bien à considérer que toutes les probabilités sont conditionnelles, c'est au sens où elles sont des probabilités conditionnelles *fondamentales*. Or, la question de savoir si les probabilités conditionnelles fondamentales sont des probabilités conditionnelles au sens du calcul ne se pose pas. D'une part, et ainsi que nous venons de le montrer, il est clair qu'elles n'en sont pas. D'autre part, la question du rapport entre causalité et probabilités telle que nous l'envisageons dans le présent travail est clairement la question du rapport entre les relations de cause à effet et les probabilités conditionnelles *d'événements*. Il en découle deux choses : d'abord que la version du paradoxe de Humphreys que nous venons de présenter n'est pas un problème qui doit nous arrêter ici ; ensuite qu'il convient que nous nous intéressions aux probabilités conditionnelles d'événements. Nous le faisons dans la prochaine sous-section, qui est plus précisément consacrée à montrer comment le paradoxe de Humphreys resurgit dans ce cadre.

6.1.2 Le paradoxe de Humphreys pour les probabilités conditionnelles d'événements

Tel qu'il est présenté dans Humphreys (1985), le paradoxe de Humphreys porte sur les probabilités conditionnelles d'événements (et non sur les probabilités conditionnelles fondamentales). Aussi la sous-section qui commence est-elle d'abord consacrée à présenter l'argument qui est développé dans Humphreys (1985). De cet argument, nous présentons la version formelle. Cela permet une exposition rigoureuse et fidèle et, à la fois, conduit à définir

[7] Gillies (2000a) p. 132.
[8] Gillies (2000a) p. 132.

un cadre auquel peut être intégrée la plupart des versions du paradoxe pour les probabilités conditionnelles d'événements. C'est ce que nous faisons à la fin de la sous-section – et à la suite de Humphreys (2004).

6.1.2.1 Un exemple

La version formelle de l'argument de Humphreys est introduite au moyen d'un exemple :

> Prenez, alors, le cas d'un phénomène physique bien connu, la transmission et la réflexion de photons par un miroir à moitié argenté. Une source de photons émis spontanément autorise les particules à rencontrer le miroir, mais le système est conçu de telle façon que tous les photons émis par la source ne frappent pas le miroir. Il est par ailleurs suffisamment isolé pour que seuls les facteurs explicitement mentionnés ici soient pertinents. Soit I_{t2} l'événement constitué par la rencontre d'un photon avec le miroir à l'instant $t2$, et soit T_{t3} l'événement constitué par l'émission d'un photon à travers le miroir à l'instant $t3$ postérieur à $t2$. Maintenant, considérez la propension conditionnelle d'événements singuliers $Pr_{t1}(.|.)$ où $t1$ est antérieur à $t2$, et prenez les assignations de valeurs de propensions suivantes :
> i) $Pr_{t1}(T_{t3}|I_{t2}B_{t1}) = p > 0$
> ii) $1 > Pr_{t1}(I_{t2}|B_{t1}) = q > 0$
> iii) $Pr_{t1}(T_{t3}|\overline{I_{t2}}B_{t1}) = 0$
> où, pour éviter les problèmes liés à la spécificité maximale, chaque propension est conditionnalisée par un ensemble complet de conditions d'arrière-plan B_{t1} qui inclut le fait qu'un photon a été émis par la source à $t0$, qui n'est pas postérieur à $t1$.[9]

Avant d'aller plus loin, nous voudrions formuler deux remarques concernant ce passage. D'abord, Humphreys parle de « propension ... d'événements singuliers » pour désigner la fonction de probabilités Pr en tant qu'elle reçoit une interprétation propensionniste. Dans un premier temps, nous préférons précisément nous en tenir à la notion de probabilité d'événements singuliers interprétée en termes propensionnistes. L'expression est certes plus lourde, mais elle est aussi plus précise. Par ailleurs, nous considérons que les probabilités dont il est question sont les probabilités en $t1$ qu'*un* photon *particulier* émis en $t0$ fasse telle ou telle chose en $t2$ ou en $t3$ (étant donné l'ensemble de conditions d'arrière-plan défini en $t1$).

Relativement à l'exemple qu'il introduit, la question posée par Humphreys est la suivante : quelle est la valeur de la probabilité $Pr_{t1}(I_{t2}|T_{t3}B_{t1})$? On remarquera que cette probabilité est d'un type particulier : il s'agit d'une

[9] Humphreys (1985) p. 561.

probabilité conditionnelle d'événement[10] telle que l'événement conditionnant est postérieur à l'événement conditionné. En d'autres termes, il s'agit d'une *probabilité conditionnelle inverse*.

6.1.2.2 Le principe (CI)

En réponse à la question de l'évaluation de $Pr_{t1}(I_{t2}|T_{t3}B_{t1})$, Humphreys défend la double égalité suivante :

$$Pr_{t1}(I_{t2}|T_{t3}B_{t1}) = Pr_{t1}(I_{t2}|\overline{T_{t3}}B_{t1}) = Pr_{t1}(I_{t2}|B_{t1}).^{11} \tag{6.1}$$

Ces égalités découlent du principe (CI) proposé par Humphreys :

Principe 6.1 (CI) *Soit B_{t1} un ensemble de conditions physiques, et A_{t2} et C_{t3} deux événements que B_{t1} sont susceptibles d'engendrer. Supposons que $Pr(C_{t3}B_{t1}) \neq 0$, $Pr(\overline{C_{t3}}B_{t1}) \neq 0$ et $Pr(B_{t1}) \neq 0$.*
Si $t1 < t2 < t3$, alors $Pr_{t1}(A_{t2}|C_{t3}B_{t1}) = Pr_{t1}(A_{t2}|\overline{C_{t3}}B_{t1}) = Pr_{t1}(C_{t2}|B_{t1})$.

Notons que ce principe ne vaut que pour l'évaluation des probabilités conditionnelles inverses telles que l'ensemble de conditions physiques de référence est antérieur au conditionnant et au conditionné de la probabilité conditionnelle. C'est dans ce cas qu'apparaît le paradoxe de Humphreys, et donc relativement à lui que se pose la question de l'interprétation propensionniste des probabilités conditionnelles. C'est toujours ainsi que nous l'entendons dans le présent chapitre. Les autres cas ne sont pas problématiques et peuvent être traités conformément à la proposition de Miller.[12]

Mais revenons à Humphreys (1985). En faveur de la double égalité 6.1 et, au-delà, en faveur du principe 6.1, Humphreys avance l'argument suivant :

> ... la propension que la particule frappe le miroir n'est pas affectée par la transmission ou non de la particule (*by whether the particule is transmitted or not*).[13]

Autrement dit, le principe (CI) découle de ce qu'un événement postérieur ne peut pas agir sur la propension qui tend à réaliser un événement antérieur.

Plus loin dans l'article, Humphreys affine et renforce sa position. Il envisage alors deux altérations possibles du système considéré : dans la première,

[10]Bien sûr, et comme toujours dans le contexte propensinniste, il s'agit aussi d'une probabilité conditionnelle fondamentale, dont le conditionnant est B_{t1}. Toutefois, pour des raisons que nous avons dites plus haut, ce n'est pas cet aspect qui retient notre attention ici.

[11]Humphreys (1985) p. 561.

[12]Miller (2002) section 4.

[13]Humphreys (1985) p. 561.

on utilise un miroir opaque et $Pr_{t1}(T_{t3}|I_{t2}B_{t1}) = 0$; dans la seconde, on utilise un miroir transparent et $Pr_{t1}(T_{t3}|I_{t2}B_{t1}) = 1$. Or, remarque Humphreys, il n'y a pas de raison que penser que ces altérations modifient $Pr_{t1}(I_{t2}|B_{t1})$. Plus généralement, la probabilité de I_{t2} n'est pas affectée par ce qui influence causalement l'occurrence ou la non-occurrence de T_{t3}, et « les événements T_{t3} et $\overline{T_{t3}}$ ne sont pas pertinents relativement à la propension pour I_{t2} »[14]. Humphreys en déduit que « ils [T_{t3} et $\overline{T_{t3}}$] peuvent être omis des facteurs sur lesquels la propension est conditionnalisée sans que sa valeur en soit affectée »[15]. De façon équivalente, conjoindre l'un ou l'autre d'entre eux au conditionnant B_{t1} n'a pas d'effet sur la valeur de la probabilité conditionnelle dont I_{t2} est le conditionné. Dès lors, la double égalité $Pr_{t1}(I_{t2}|T_{t3}B_{t1}) = Pr_{t1}(I_{t2}|\overline{T_{t3}}B_{t1}) = Pr_{t1}(I_{t2}|B_{t1})$ est bien satisfaite.

6.1.2.3 Le paradoxe

Une fois le principe (CI) dûment justifié, Humphreys peut établir le résultat qui fait paradoxe. Ce résultat est un résultat d'incomptatibilité entre (CI) et les propriétés des probabilités conditionnelles. Plus exactement, Humphreys dérive deux contradictions. La première dérivation a pour prémisses : les propositions i) à iii) introduites à l'occasion de la présentation de l'exemple (et qui sont peu discutables), le principe (CI) et le théorème des probabilités totales pour les événements binaires :

Théorème 6.1 (Probabilités totales pour les événements binaires) *Soit A, B, C trois événements tels que* $P(BC) \neq 0$, $P(\overline{B}C) \neq 0$ *et* $P(C) \neq 0$. *Alors :*
$P(A|C) = P(A|BC).P(B|C) + P(A|\overline{B}C).P(\overline{B}|C)$.

La seconde des dérivations de Humphreys a pour prémisses, les propositions i) à iii), le principe (CI) et le théorème de Bayes pour les événements binaires :

Théorème 6.2 (Bayes pour les événements binaires) *Soit A, B, C trois événements tels que* $P(C) \neq 0$, $P(BC) \neq 0$ *et* $P(\overline{B}C) \neq 0$. *Alors :*
$P(B|AC) = P(A|BC).P(B|C)/[P(A|BC).P(B|C) + P(A|\overline{B}C).P(\overline{B}|C)]$.

De deux résultats d'incompatibilité que nous venons de décrire, Humphreys conclut que le propensionnisme ne constitue pas une interprétation admissible (au sens de la sous-section 5.1.3) du calcul des probabilités étendu aux probabilités conditionnelles. Selon Humphreys, il en découle que le calcul des probabilités n'est pas une théorie qui décrit convenablement l'ensemble des phénomènes aléatoires.[16]

[14] Humphreys (1985) p. 563.
[15] Humphreys (1985) p. 563.
[16] Humphreys (1985) pp. 569-570 et Humphreys (2004) p. 679.

6.1.2.4 Le paradoxe de Humphreys généralisé

A la rigueur, Humphreys (1985) offre une seule issue à qui veut dissoudre le paradoxe : montrer que (CI) n'est pas le principe qui doit présider à l'évaluation des probabilités conditionnelles inverses dans un cadre propensionniste. En fait, les opposants à Humphreys (1985) ne s'en prennent pas au seul (CI) et, surtout, distinguent rarement entre la question de l'évaluation des probabilités conditionnelles inverses et celle de l'interprétation propensionniste des probabilités conditionnelles. La généralisation du paradoxe opposée à l'ensemble de ces réponses dans Humphreys (2004) prend en compte cette situation.

Dans cet article, Humphreys commence par identifier dans la littérature successive à Humphreys (1985) deux principes rivaux de (CI) pour l'évaluation des probabilités conditionnelles inverses :

– le principe d'influence nulle :

Principe 6.2 (ZI) *Soit B_{t1} un ensemble de conditions physiques, et A_{t2} et C_{t3} deux événements que B_{t1} sont susceptibles d'engendrer. Supposons que $Pr(C_{t3}B_{t1}) \neq 0$, $Pr(\overline{C_{t3}}B_{t1}) \neq 0$ et $Pr(B_{t1}) \neq 0$.*
Si $t1 < t2 < t3$, alors $Pr_{t1}(A_{t2}|C_{t3}B_{t1}) = 0$.
Ce principe exprime l'idée selon laquelle C_{t3} ne peut pas avoir d'influence, *via* une propension, sur A_{t2} si $t3$ est postérieur $t2$. Il est défendu en particulier dans Fetzer (1981) ;

– le principe de fixité :

Principe 6.3 (FP) *Soit B_{t1} un ensemble de conditions physiques, et A_{t2} et C_{t3} deux événements que B_{t1} sont susceptibles d'engendrer. Supposons que $Pr(C_{t3}B_{t1}) \neq 0$, $Pr(\overline{C_{t3}}B_{t1}) \neq 0$ et $Pr(B_{t1}) \neq 0$.*
Si $t1 < t2 < t3$, alors $Pr_{t1}(A_{t2}|A_{t3}B_{t1}) = 0$ ou 1.
Ce principe exprime l'idée suivante : si $t3$ est postérieur à $t2$, alors, à l'instant $t3$ qui caractérise C_{t3}, soit A_{t2} est advenu – et alors sa probabilité vaut 1, soit A_{t2} n'est pas advenu – et alors sa probabilité vaut 0. Cette position est défendue en particulier dans Milne (1986).[17]

Une fois identifiés ces deux principes rivaux de (CI), Humphreys montre que la contradiction mise en évidence dans Humphreys (1985) se dérive si l'on adopte l'un ou l'autre de ces deux principes aussi bien que si l'on s'en tient à (CI).[18] Finalement, Humphreys établit une typologie des propositions existantes pour l'interprétation propensionniste des probabilités conditionnelles et montre que chaque type de propositions conduit à évaluer les probabilités conditionnelles inverses conformément à l'un des trois principes (CI), (ZI) et (FP). Le paradoxe de Humphreys se trouve alors généralisé. On notera

[17] Milne (1986) p. 130–131.
[18] Humphreys (2004) p. 670.

toutefois que cette généralisation prend en compte les seules propositions *existantes* pour l'interprétation propensionniste des probabilités conditionnelles et l'évaluation des probabilités conditionnelles inverses. Pour le dire autrement, Humphreys ne propose par d'objection qui vaudrait en principe contre *toutes* les propositions possibles.

Nous n'avons pas décrit, et ne décrirons pas l'ensemble des quatre types d'interprétations propensionnistes des probabilités conditionnelles que Humphreys identifie. Positivement, un seul de ces quatre types retiendra notre attention ici. Il s'agit du type des interprétations qu'Humphreys appelle « de co-production ». Nous soutenons qu'elles n'ont pas exactement le même statut que les autres interprétations propensionnistes des probabilités conditionnelles, et nous nous attachons à le montrer maintenant.

6.1.3 Spécificité des interprétations de co-production

Des interprétations de co-production des probabilités conditionnelles sont défendues en particulier par Miller[19] et par McCurdy[20]. Humphreys les caractérise dans les termes suivants :

> Une *interprétation de co-production* considère que la propension conditionnelle est localisée dans les conditions structurelles qui sont présentes à l'instant initial t, avec $Pr_t(.|.)$ la propension à produire les événements qui constituent les deux arguments de la propension conditionnelle.[21]

Nous ne nous arrêtons pas maintenant à cette caractérisation, sur laquelle nous reviendrons longuement dans la suite du chapitre. Plutôt, nous nous en tenons ici à défendre la thèse selon laquelle les interprétations de co-production ont un statut particulier dans le champ des interprétations propensionnistes des probabilités conditionnelles. En faveur de cette thèse, nous avançons trois arguments.

Le premier de ces arguments consiste simplement à remarquer que Humphreys lui-même discute les interprétations de co-production beaucoup plus longuement que les autres types d'interprétations propensionnistes des probabilités conditionnelles qu'il identifie.

Notre deuxième argument procède d'une tentative pour expliquer l'inégalité de traitement que nous venons de remarquer. Plus précisément, l'argument consiste à faire valoir que les tenants d'une interprétation de

[19] Miller (1994) section 9.5.
[20] McCurdy (1996).
[21] Humphreys (2004) p. 671.

co-production sont les seuls, parmi les auteurs qui ont proposé des interprétations propensionnistes des probabilités conditionnelles, à ne pas revendiquer l'un des trois principes (CI), (ZI) et (FP) pour l'évaluation des probabilités conditionnelles inverses. Cela explique bien que Humphreys leur consacre un passage plus long qu'aux autres interprétations qu'il envisage : en vue de généraliser le paradoxe, Humphreys doit commencer par montrer que les interprétations de co-production conduisent à évaluer les probabilités conditionnelles inverses conformément à l'un des trois principes qu'il a identifiés. En outre, cela implique que, parmi les opposants au paradoxe de Humphreys, les tenants d'une interprétation de co-production sont les seuls à proposer une défense qui, si elle est valide, rend Humphreys (2004) non concluant. Ainsi, les interprétations de co-production se présentent comme la seule voie ouverte pour une résolution du paradoxe de Humphreys.

Enfin, si la stratégie officielle de Humphreys consiste à montrer que les interprétations de co-production conduisent au principe (CI) pour l'évaluation des probabilités conditionnelles inverses, elle laisse finalement la place au véritable reproche de Humphreys à l'encontre des interprétations de co-production :

> l'interprétation de co-production elle-même est sérieusement défectueuse en tant qu'interprétation des propensions conditionnelles...[22]

et :

> ...on peut faire peser contre cette conception les arguments présentés dans la section 6 ci-dessus et selon lesquels on a conservé la structure des probabilités conditionnelles au prix de retirer aux propensions conditionnelles ce qui traditionnellement reconnu comme leur étant essentiel.[23]

Il apparaît alors que la principale tare des interprétations de co-production aux yeux de Humphreys est le fait qu'elles ne sont pas fidèles à l'esprit du propensionnisme, et donc pas acceptables pour un propensionniste. La question n'est plus alors seulement celle de l'évaluation des probabilités conditionnelles inverses. Elle est aussi la question de l'interprétation propensionniste des probabilités conditionnelles. Le débat se porte ainsi sur le terrain de ce qui engendre le paradoxe de Humphreys.

Les trois arguments que nous venons de présenter ne montrent pas seulement que les interprétations de co-production ont un statut différent de celui des autres interprétations propensionnistes des probabilités conditionnelles.

[22]Humphreys (2004) p. 673.
[23]Humphreys (2004) p. 677.

Ils tendent également à établir que s'intéresser aux interprétations de co-production doit permettre de mettre le doigt sur le point où se noue le paradoxe de Humphreys et, peut-être, de le résoudre. Dans ces conditions, le projet d'analyse du rapport entre probabilités conditionnelles et causalité dans le contexte propensionniste nous porte à une analyse méthodique du désaccord qui oppose Humphreys aux tenants d'interprétations de co-production. Cette analyse est menée dans la prochaine section.

6.2 Analyse d'un désaccord

Nous avons indiqué déjà que l'interprétation de co-production a deux défenseurs principaux : David Miller et Christopher McCurdy. Toutefois, l'analyse que nous menons dans la section qui commence prend en compte le seul McCurdy (1996). La première raison en est que ce texte porte le plus directement sur l'argument formel que nous avons présenté dans la sous-section 6.1.2. En outre, McCurdy (1996) est le texte sur lequel porte la charge de Humphreys dans Humphreys (2004). En effet, pour ce qui est de Miller, Humphreys accepte ses arguments contre les versions du paradoxe correspondant aux principes (ZI) et (FP), et pour le reste renvoie à la critique adressée à McCurdy. Enfin, il n'est pas contestable que McCurdy (1996) est topique dans le champ des interprétations de co-production des probabilités conditionnelles. Ainsi Miller écrit-il : « McCurdy [1996], un papier avec lequel je suis largement en accord (même s'il est suggéré à tort à la p. 106 que je considère que les propensions sont fondamentalement des propensions à engendrer des fréquences) »[24].

Avant d'en venir à la dispute entre Humphreys et McCurdy, il convient de décrire le socle à partir duquel elle se déploie. Ce socle se compose essentiellement de deux thèses : d'une part la thèse du bien-fondé du propensionnisme de cas singuliers, d'autre part celle du bien-fondé du débat sur les probabilités conditionnelles. Plus précisément, les deux auteurs considèrent que :

1. il y a un sens à parler de la probabilité physique d'un événement singulier ;
2. le propensionnisme vise à interpréter ces probabilités ;
3. la notion d'interprétation propensionniste des probabilités conditionnelles inverses n'est pas dépourvue de sens. En cela, ils se distinguent en particulier de Salmon[25].

[24] Miller (2002) Introduction.
[25] Salmon (1979).

Pour ce qui est, ensuite, du propensionnisme de cas singuliers lui-même, chacun des deux auteurs distingue explicitement entre le système physique dont les propensions sont des propriétés et les événements que ces propensions tendent à réaliser.[26]

Cela étant posé, nous pouvons en venir à l'analyse du désaccord entre nos deux auteurs. Dans la première sous-section, nous nous concentrons sur l'exemple introduit dans Humphreys (1985) – ou plus exactement sur la différence des analyses que Humphreys et McCurdy en proposent respectivement. La seconde sous-section vise à mettre au jour les termes plus fondamentaux du désaccord entre les deux auteurs.

6.2.1 Le désaccord sur l'exemple introduit par Humphreys

Dans McCurdy (1996), l'analyse de Humphreys est attaquée de manière frontale. Plus précisément, McCurdy critique l'analyse par Humphreys de l'exemple de la transmission de photons à partir duquel se déploie l'« argument formel » que nous avons présenté dans la sous-section 6.1.2. C'est cette critique que nous présentons maintenant.

6.2.1.1 L'objection de McCurdy à l'analyse de Humphreys

La critique de McCurdy porte d'abord sur les altérations du système initial que Humphreys envisage – ou plus exactement sur ce que Humphreys prétend tirer de l'examen de ces altérations. Rappelons que les altérations envisagées consistent à remplacer le miroir initial, semi-opaque, par un miroir opaque d'abord et par un miroir transparent ensuite. Humphreys constate que ces altérations laissent inchangée la probabilité $Pr_{t1}(I_{t2}|B_{t1})$ que le photon frappe le miroir à l'instant $t2$, en déduit que l'occurrence ou la non-occurrence de T_3 n'a pas d'influence causale sur celle de I_{t2}, puis conclut à la vérité du principe (CI).

Au raisonnement dont nous venons de retracer les grandes lignes, McCurdy objecte ceci[27] : les altérations considérées montrent seulement que des facteurs qui influencent causalement T_{t3} et qui agissent *après que le photon a frappé le miroir* n'ont pas d'influence sur la propension qui tend à faire advenir I_{t2}. En termes probabilistes, les altérations considérées par Humphreys montrent que la valeur de $Pr_{t1}(I_{t2}|B_{t1})$ ne dépend ni de celle de $Pr_{t1}(T_{t3}|I_{t2}B_{t1})$, ni de celle de $Pr_{t1}(\overline{T_{t3}}|I_{t2}B_{t1})$. Mais elles ne nous disent rien

[26] Humphreys (2004) p. 669 et McCurdy (1996) pp. 107–108.

[27] McCurdy (1996) pp. 114–115.

sur les probabilités $Pr_{t1}(I_{t2}|T_{t3}B_{t1})$ et $Pr_{t1}(I_{t2}|\overline{T_{t3}}B_{t1})$ – qui nous intéressent. Ces probabilités peuvent très bien avoir une valeur différente de celle de $Pr_{t1}(I_{t2}|B_{t1})$, en particulier s'il existe des facteurs agissant *entre* $t0$ *et* $t2$ et qui influencent causalement à la fois I_{t2} et T_{t3}.

Or, poursuit McCurdy, le système considéré est précisément tel que des facteurs de ce type existent :

> l'arrangement de transmission de photons lui-même (tel qu'il est décrit par B_{t1}) fournit une foule de facteurs causaux communs. Ce fait est responsable de l'échec du principe (CI) : si le système produit l'événement T_{t3}, alors il a dû exhiber certains facteurs causaux, dont certains ont une influence sur l'événement I_{t2}. Etant donné que la fonction de propension est définie *pour* un système, ces influences sont prises en compte par l'assignation de valeurs aux propensions. La méthode de Humphreys pour justifier (CI) sur la base des relations entre les événements singuliers I_{t2}, T_{t3} et $\overline{T_{t3}}$ ne prend pas en compte ces influences.[28]

Selon McCurdy, prendre en compte ces influences conduit à reconnaître que conditionnaliser sur T_{t3} affecte la probabilité $Pr_{t1}(I_{t2}|B_{t1})$, et rejeter (CI).

Positivement, McCurdy soutient que la probabilité conditionnelle inverse examinée vaut 1. En effet, ce système est par hypothèse tel que $Pr_{t1}(T_{t3}|\overline{I_{t2}}B_{t1}) = 0$ – autrement dit : tel que le photon ne peut pas être transmis s'il n'a pas frappé le miroir. De façon équivalente :

> ...si le système produit un photon qui est transmis en $t3$, alors le système doit aussi produire un photon qui frappe la miroir en $t2$.[29]

La thèse soutenue par McCurdy est donc finalement la suivante : les propriétés physiques du dispositif envisagé impliquent que $Pr_{t1}(I_{t2}|T_{t3}B_{t1}) = 1$.

6.2.1.2 La réponse de Humphreys à McCurdy

Dans Humphreys (2004), Humphreys revient sur l'exemple du système de transmission de photons, et maintient que la probabilité conditionnelle inverse considérée doit être évaluée conformément à (CI). Contre McCurdy, il nie l'existence des facteurs causaux communs à I_{t2} et T_{t3} agissant entre $t0$ et $t2$. Plus précisément, il soutient que leur existence est une illusion engendrée par des caractéristiques contingentes de l'exemple considéré, « des aspects quasi-déterministes de la propension en $t1$ $Pr_{t1}(I_{t2}|B_{t1})$, qui est fondamentalement indéterministe »[30].

[28] McCurdy (1996) p. 116.
[29] McCurdy (1996) pp. 110–111.
[30] Humphreys (2004) p. 674.

Pour dissiper l'illusion, Humphreys introduit un exemple « formellement identique à l'exemple du photon »[31], mais ne présentant pas ses « aspects quasi-déterministes ». L'indéterminisme indiscutable du nouvel exemple est assuré par le fait que l'équivalent de I_{t2} est un épisode de désintégration radioactive. Selon Humphreys, pour ce nouvel exemple :

> il devrait être clair que ... il n'existe pas de facteurs causaux communs entre $t1$ et $t2$ sur la base desquels on pourrait affirmer sans erreur que si le système a produit l'événement D_{t3} [l'équivalent de T_{t3}], alors il a dû exhiber certains facteurs causaux entre $t1$ et $t2$ dont certains ont une influence sur E_{t2} [l'équivalent de I_{t2}]. Le principe (CI) est donc vrai et, il me semble, manifestement vrai.[32]

Humphreys considère que ces conclusions, si elles sont plus facilement aperçues quand on considère le nouveau système, valaient déjà du système envisagé dans Humphreys (1985).

6.2.1.3 Analyse de la réponse de Humphreys à McCurdy

Il nous semble que l'argument proposé par Humphreys souffre de deux faiblesses. D'une part, l'analyse du rapport entre les deux exemples nous paraît insuffisante à justifier que les conclusions établies pour le second valent *ipso facto* pour le premier. En particulier, rien ne vient garantir que la différence entre les « aspects quasi-déterministes » du système initial et « la nature irréductiblement indéterministe »[33] du second n'a pas pour corrélat des principes différents pour l'évaluation des probabilités conditionnelles inverses. D'autre part, et en amont de la difficulté que nous venons d'identifier, la justification de la thèse selon laquelle le principe (CI) doit présider à l'évaluation des probabilités conditionnelles inverses dans le nouveau système n'est pas satisfaisante. Plus précisément, il nous semble qu'elle est insatisfaisante à deux titres.

En premier lieu, la thèse de l'absence de facteurs causaux communs au conditionné et au conditionnant de la probabilité considérée est insuffisamment établie. Humphreys considère que cette absence est garantie par « la nature irréductiblement indéterministe » du système envisagé. Plus précisément, le raisonnement de Humphreys semble être le suivant : puisque E_{t2} est un épisode de désintégration radioactive, il n'a pas de cause, et *a fortiori* pas de cause qu'il partagerait avec D_{t3}. Mais peut-on ainsi assimiler l'absence de condition suffisante propre au phénomène de désintégration radioactive et l'absence de cause ? En particulier, ne peut-on pas considérer que

[31] Humphreys (2004) p. 674.
[32] Humphreys (2004) pp. 673–674.
[33] Humphreys (2004) p. 675.

l'arrangement du système envisagé est lui-même une cause de E_{t2} s'il advient, dans la mesure en particulier où il en augmente considérablement la probabilité toutes choses étant égales par ailleurs ? McCurdy semble prendre en compte une telle possibilité quand il affirme que « l'arrangement . . . lui-même fournit une foule de facteurs causaux communs »[34]. En ignorant ce point, il nous semble que Humphreys échoue à établir qu'il n'existe pas de facteurs causaux communs à E_{t2} et D_{t3} – et du coup à établir que l'argument de McCurdy contre (CI) disparaît quand on dissipe l'illusion de déterminisme.

En second lieu, la justification de (CI) dans Humphreys (2004) consiste tout entière à critiquer la thèse de McCurdy selon laquelle il existerait des facteurs causaux communs à I_{t2} et T_{t3} agissant entre $t1$ et $t2$. Ainsi Humphreys ne donne-t-il pas de nouveaux arguments positifs en faveur de (CI). Il ne fait que réaffirmer, sans la justifier, la thèse exposée dans Humphreys (1985) : « le principe (CI) est [. . .] vrai et, il me semble, manifestement vrai pour ce système »[35]. En outre, il ne s'attaque pas à l'idée de McCurdy selon laquelle les propriétés physiques du système considéré seraient telles que la probabilité discutée vaut 1.

En définitive, nous ne voyons pas dans Humphreys (2004) de nouvelles bonnes raisons de penser que (CI) est le principe qui doit présider à l'évaluation des probabilités conditionnelles inverses. De l'autre côté, il nous semble que l'argument de McCurdy en faveur de la valeur 1 pour la probabilité $Pr_{t1}(I_{t2}|T_{t3}B_{t1})$ est, finalement, insensible aux « aspects quasi-déterministes » du système initialement considéré et qu'il vaut, *mutatis mutandis*, du système de « nature irréductiblement indéterministe ». En effet, de même qu'il écrivait du système initial que « si le système produit un photon qui est transmis en $t3$, alors le système doit aussi produire un photon qui frappe la miroir en $t2$ »[36], McCurdy pourrait dire du nouveau système que s'il produit une particule alpha détectée en $t3$, il doit aussi produire une émission de particule alpha en $t2$.

Dans ces conditions, nous soutenons que le débat entre Humphreys et McCurdy est inchangé par Humphreys (2004). Chacun des deux auteurs campe sur ses positions, et nous y voyons un indice de ce que la divergence entre les deux auteurs s'étend bien au-delà de la question de savoir comment doit être évaluée la probabilité $Pr_{t1}(I_{t2}|T_{t3}B_{t1})$ dans l'exemple du système de transmission de photons.

[34]McCurdy (1996) p. 116.
[35]Humphreys (2004) p. 675.
[36]McCurdy (1996) pp. 110–111.

6.2.2 Le désaccord au-delà de l'évaluation de $Pr_{t1}(I_{t2}|T_{t3}B_{t1})$

Pour commencer, notons que le désaccord relatif à l'évaluation de $Pr_{t1}(I_{t2}|T_{t3}B_{t1})$ porte en fait sur l'évaluation des probabilités conditionnelles inverses *de manière générale*. D'un côté, nous avons vu (dans le paragraphe 6.1.2.2) que Humphreys soutient que toutes ces probabilités doivent être évaluées conformément au principe (CI). De l'autre, McCurdy (comme Miller d'ailleurs) semble admettre qu'il n'existe pas de formule générale de la détermination de leur valeur par l'ensemble de conditions physiques qu'on considère – et qu'il faut donc toujours en revenir à lui pour les évaluer. C'est d'ailleurs ce qu'il fait pour la probabilité $Pr_{t1}(I_{t2}|T_{t3}B_{t1})$ du système de transmission de photons. Dans la sous-section qui commence, nous analysons plus précisément ce qui fonde l'une et l'autre de ces positions. Cette analyse fait apparaître que le désaccord porte en fait sur l'interprétation des probabilités conditionnelles et, au-delà encore, sur le statut du propensionnisme.

6.2.2.1 Interprétation des probabilités conditionnelles

La position de Humphreys. Pour ce qui est, d'abord, de Humphreys, rappelons que son argument en faveur de (CI) comme principe général pour l'évaluation des probabilités conditionnelles inverses est le suivant : un conditionnant postérieur au conditionné ne peut pas avoir d'influence sur la propension qui tend à réaliser ce conditionné – et ne saurait donc en modifier la valeur. Cette justification suppose que la valeur d'une probabilité conditionnelle $P_{t1}(A_{t2}|C_{t3})$[37] diffère de celle de la probabilité absolue $P_{t1}(A_{t2})$ seulement si l'occurrence de C en $t3$ modifie *physiquement* la propension qui tend à réaliser l'événement A_{t2}.

Il apparaît alors que la probabilité conditionnelle $P_{t1}(A_{t2}|C_{t3})$ mesure la propension en $t1$ du système considéré à produire A_{t2} *en tant qu'elle est éventuellement modifiée par l'occurrence de* C_{t3}. Dans le cas où $t3$ est antérieur à $t2$, l'occurrence de C_{t3} peut effectivement venir modifier la propension qui tend à réaliser A_{t2} ; $P_{t1}(A_{t2}|C_{t3})$ mesure alors la propension qui tend à réaliser à A_{t2} à l'instant $t3$ de l'occurrence de C_{t3} : $P_{t1}(A_{t2}|C_{t3}) = P_{t3}(A_{t2})$. Dans le cas où $t3$ est postérieur à $t2$, l'occurrence de C_{t3} ne peut pas modifier la propension qui tend à réaliser A_{t2} ; la valeur de cette propension est

[37] Dans ce paragraphe, nous faisons l'économie d'une référence explicite à l'ensemble de conditions d'arrière-plan relativement auquel une probabilité est définie – *i.e.* au conditionnant fondamental. En effet, les problèmes soulevés de manière spécifique par cet aspect du propensionnisme ne sont abordés que plus loin dans le texte.

donc inchangée : conformément à (CI), $P_{t1}(A_{t2}|C_{t3}) = P_{t1}(A_{t2})$.[38] De façon plus générale, conditionaliser revient pour Humphreys à prendre en compte l'influence (physique) possible de l'occurrence du conditionnant sur la propension qui tend à réaliser le conditionné.

La position de McCurdy. De son côté et contre (CI), McCurdy considère qu'une probabilité conditionnelle inverse peut avoir une valeur différente de celle de la probabilité absolue du conditionné. Cela suppose que la différence entre la probabilité conditionnelle $P_{t1}(A_{t2}|C_{t3})$ et la probabilité absolue $P_{t1}(A_{t2})$ ne correspond pas à l'influence possible de l'occurrence du conditionnant C_{t3} sur la propension qui tend à réaliser le conditionné A_{t2}. Positivement, la différence entre $P_{t1}(A_{t2}|C_{t3})$ et $P_{t1}(A_{t2})$ renvoie à ce que l'occurrence de C_{t3} *implique* relativement à l'ensemble des conditions d'arrière-plan.

On aura compris que la relation d'implication dont il est question ici n'est pas de nature causale, ni même plus généralement de nature physique. Il convient de souligner qu'elle n'est pas non plus (ou en tout cas pas d'abord) de nature épistémique : on ne s'intéresse à ce que l'occurrence de C_{t3} permet d'inférer relativement à l'ensemble de conditions physiques d'arrière-plan.[39] Positivement, il semble que cette relation doit plutôt être caractérisée comme logique, au sens précis où elle a pour second terme les propriétés que l'ensemble de conditions d'arrière-plan ne peut pas ne pas exhiber s'il en vient à produire l'événement conditionnant C_{t3}. Ces propriétés sont quant à elles de nature physique ; elles viennent définir plus précisément l'ensemble de conditions d'arrière-plan.

Dans ces conditions, la conditionalisation est interprétée comme une respécification de l'ensemble de conditions d'arrière-plan : conditionaliser sur C_{t3}, c'est exactement substituer à la description initiale de l'ensemble de conditions physiques d'arrière-plan la description plus précise qui découle logiquement du fait que cet ensemble produit le conditionnant. Les probabilités conditionnelles mesurent les propensions relatives à cet ensemble redéfini, et non plus à l'ensemble initial. Dans ces conditions, que l'événement conditionnant soit antérieur ou postérieur à l'événement conditionné ne fait pas

[38] Selon cette lecture, Humphreys souscrit à une interprétation « d'évolution temporelle » des probabilités conditionnelles (Humphreys (2004) p. 672). Cette analyse s'accorde avec le fait que Humphreys considère que les interprétations d'évolution temporelle conduisent à évaluer les probabilités conditionnelles inverses conformément à (CI) (Humphreys (2004) pp. 672 et 677).

[39] Une remarque du même ordre est déjà introduite par Humphreys : « ... McCurdy ne commet pas l'erreur d'en appeler au fait que nous pouvons inférer avec certitude de la structure de l'arrangement expérimental que quand un photon a été transmis il a dû frapper le miroir » (Humphreys (2004) p. 674).

de différence remarquable, et McCurdy a raison de soutenir que les probabilités conditionnelles inverses ne posent pas de problème spécifique.

Dans le paragraphe qui s'achève, nous avons montré comment la dispute relative à l'évaluation d'une probabilité conditionnelle inverse s'alimente à un désaccord plus profond, relatif à la façon dont les probabilités conditionnelles doivent être interprétées dans un cadre propensionniste. Il nous semble possible d'aller plus loin, et de montrer que le désaccord sur l'interprétation des probabilités conditionnelles est corrélatif d'une divergence de vues concernant le statut même du propensionnisme. C'est ce que nous nous proposons de faire dans le prochain paragraphe.

6.2.2.2 Statut du propensionnisme

La divergence entre Humphreys et McCurdy relativement au statut du propensionnisme se manifeste relativement à la question de savoir à combien de relations les probabilités conditionnelles font référence. Nous verrons à la fin du paragraphe qui commence qu'elle implique une divergence relativement au statut des interprétations proposées pour la conditionalisation.

Une relation ou deux relations ? Dans le dernier paragraphe, nous avons montré que Humphreys considère qu'une probabilité conditionnelle mesure la propension qui tend à réaliser le conditionné, en tant qu'elle est éventuellement modifiée par l'occurrence du conditionnant. L'interprétation des probabilités conditionnelles semble alors engager une référence non plus à une, mais à *deux* relations de nature causale : d'abord la propension à réaliser l'événement conditionné, ensuite la relation en vertu de laquelle l'occurrence du conditionnant peut éventuellement modifier cette propension. L'interprétation de la conditionalisation n'a pas les mêmes réprecussions chez McCurdy.

La divergence que nous venons de pointer a une place centrale dans la structure argumentative de Humphreys (2004). D'une part, en effet, nous avons soutenu que Humphreys échoue à montrer effectivement que l'interprétation des probabilités conditionnelles retenue par McCurdy doit le conduire à évaluer les probabilités conditionnelles inverses conformément à (CI). Du coup, son seul argument valide contre l'idée selon laquelle McCurdy aurait résolu le paradoxe de Humphreys consiste à rejeter l'interprétation des probabilités conditionnelles qui sous-tend la proposition de McCurdy. D'autre part, le rejet de l'interprétation des probabilités conditionnelles qui sous-tend la proposition de McCurdy se joue précisément sur la question des relations auxquelles réfèrent les probabilités :

> ... [L'interprétation des probabilités conditionnelles proposée par McCurdy] présente la relation entre les événements conditionnant et conditionné comme une relation entre mesures de probabilités plutôt que comme une relation matérielle entre événements concrets. Selon cette conception, il n'y pas de relation de propension entre les événements conditionnant et conditionné de la probabilité conditionnelle $P(A|B)$. Il ne s'agit donc pas d'une propension conditionnelle de cas singulier à proprement parler.[40]

En définitive, la raison pour laquelle Humphreys considère que McCurdy n'a pas résolu le paradoxe consiste tout entière dans ceci que McCurdy envisage d'interpréter la conditionalisation indépendamment de la référence à des propensions conditionnelles.

Statut du propensionnisme selon Humphreys. Maintenant, si le fait de ne pas introduire de propensions conditionnelles est inacceptable pour Humphreys, c'est en tant qu'il n'est pas vraiment propensionniste, qu'il va contre l'esprit du propensionnisme. Ce diagnostic s'alimente à la thèse selon laquelle :

> Un attrait majeur des propensions de cas singuliers a toujours été le fait qu'elles mettent l'accent non plus sur les résultats d'expériences, mais sur les dispositions physiques qui produisent ces résultats. [41]

Autrement dit, la position de Humphreys à propos des probabilités conditionnelles a pour fondement l'idée selon laquelle l'apport principal du propensionnisme à la philosophie des probabilités consiste à considérer que les probabilités singulières font référence à des dispositions physiques.

Selon cette conception, l'introduction du propensionnisme consiste à mettre en lumière un nouveau type d'objets – des dispositions indéterministes physiquement fondées – et le propensionnisme est d'abord et avant tout la théorie de ces nouveaux objets. Il n'a partie liée avec les probabilités que dans un second temps et éventuellement dans la mesure où les mesures de propensions se trouvent se comporter comme des probabilités absolues. Une fois qu'on a pris en compte les propensions, on est conduit à envisager des propensions plus complexes, dyadiques au sens où elles sont relatives à des couples d'événements. Chacune de ces propensions plus complexes correspond à la modification de la propension à réaliser un autre qui découle de l'occurrence d'un autre événement. En tant qu'elle est éventuellement modifiée par cette action, la propension initiale devient une « propension conditionnelle ». Dans

[40] Humphreys (2004) p. 675.
[41] Humphreys (2004) p. 675.

ce cadre, le paradoxe de Humphreys consiste exactement dans une réponse négative à la question de savoir si les mesures de ces propensions conditionnelles se comportent comme des probabilités conditionnelles.

Statut de l'interprétation de la conditionalisation. Selon la position que nous venons d'attribuer à Humphreys, le propensionnisme est la théorie des propensions, dont les unes sont absolues et les autres conditionnelles. Pour Humphreys, le propensionnisme considéré comme théorie des probabilités absolues *contient analytiquement* la proposition d'interprétation des probabilités conditionnelles à laquelle lui-même se range. La question de l'interprétation propensionniste des probabilités conditionnelles ne se pose donc pas comme telle. Elle se pose seulement secondairement, comme la question de savoir si les propensions conditionnelles sont des probabilités conditionnelles.

A l'inverse, la position de McCurdy se caractérise précisément par ce qu'il envisage comme telle la question de l'interprétation des probabilités conditionnelles. Pour comprendre que c'est effectivement ce qu'il fait, il convient de revenir sur le résultat établi dans Lewis (1976). Selon ce résultat, sauf dans certains cas triviaux, il n'existe pas d'objet conditionnel $C \Rightarrow A$ tel que la probabilité conditionnelle $Pr(A|C)$ a la même valeur que la probabilité absolue $Pr(C \Rightarrow A)$.[42] En d'autres termes, Lewis (1976) établit que la conditionalisation ne peut pas être interprétée comme la substitution d'un argument conditionnel à un argument absolue pour une fonction de probabilités inchangée.[43] Elle semble plutôt devoir être interprétée comme une re-définition de la fonction de probabilités, qu'on continue d'appliquer au même argument. Or, nous l'avons vu, c'est bien là précisément ce que propose McCurdy.

Finalement, l'analyse du désaccord entre Humphreys et McCurdy fait apparaître que le paradoxe de Humphreys repose sur une conception du propensionnisme comme *théorie* de ces dispositions indéterministes que sont les propensions (absolues et conditionnelles). Le propensionnisme ainsi conçu est une théorie dont l'existence est légitime : une fois qu'on a reconnu qu'il existe des propensions[44], il y a du sens à se demander quelles sont leurs propriétés –

[42]Lewis (1976) pp. 300–303.

[43]Dans ces conditions, il n'est pas surprenant que Humphreys trouve que les propensions conditionnelles ne sont pas de probabilités conditionnelles. En effet, sous la conception qu'il défend, conditionaliser revient à substituer un argument à un autre pour une fonction de probabilités inchangée – car toujours relative au même ensemble de conditions physiques.

[44]Si Humphreys reconnaît l'existence des propensions, c'est seulement en tant qu'elles sont des dispositions indéterministes fondées dans les choses. Mais il refuse l'idée selon

et en particulier si les propensions conditionnelles sont des probabilités conditionnelles. D'un autre côté, McCurdy considère que le propensionnisme tel que nous l'avons présenté dans le chapitre 5 est une *interprétation*[45] des probabilités absolues, à laquelle peut venir s'adjoindre une interprétation de la conditionalisation qui ne tombe pas sous le coup du paradoxe de Humphreys.

La position de McCurdy est tenable seulement si Humphreys a tort de considérer que le propensionnisme tel que nous l'avons présenté dans le chapitre 5 contient analytiquement une théorie des probabilités conditionnelles. Or, il nous semble que c'est bien le cas. Pour s'en convaincre, il suffit de considérer la diversité des principes d'évaluation des probabilités inverses que nous avons présentés dans la sous-section 6.1.2 – (CI), (ZI), (FP). Ces différents principes, en effet, reposent sur autant de conceptions de ce que la théorie propensionniste des probabilités absolues implique relativement aux probabilités conditionnelles. Dans ces conditions, la diversité des principes proposés pour l'évaluation des probabilités conditionnelles inverses constitue une réduction à l'absurde de la thèse selon laquelle la théorie propensionniste des probabilités absolues contiendrait une théorie des probabilités conditionnelles. McCurdy a donc raison de considérer que le propensionnisme pour les probabilités absolues laisse ouverte la question de l'interprétation des probabilités conditionnelles, et avec elle la possibilité d'une résolution du paradoxe de Humphreys.

La tâche qui s'ouvre alors consiste précisément à proposer une interprétation propensionniste des probabilités conditionnelles qui ne tombe pas sous le coup de la critique de Humphreys. Nous nous attelons à cette tâche en deux temps. Dans la prochaine section, nous proposons une interprétation propensionniste de la conditionalisation. Dans la section suivante, nous discutons la question de savoir si cette interprétation résout le paradoxe de Humphreys, c'est-à-dire si l'interprétation proposée pour les probabilités conditionnelles est admissible (au sens de la sous-section 5.1.3).

---

laquelle les propensions existeraient comme des entités autonomes, en plus des conditions physiques qui les déterminent et des événements qui les manifestent. Ainsi que nous l'avons mentionné déjà, les arguments développés dans le présent chapitre ne dépendent pas de la position qu'on adopte sur ce point.

[45] La distinction entre théorie des propensions et interprétation propensionniste des probabilités est présente chez McCurdy (McCurdy (1996) p. 119), sans y être complètement thématisée.

6.3 Proposition d'interprétation propensionniste de la conditionalisation

A ce point de notre analyse, il peut sembler que McCurdy (1996) non seulement suggère que le propensionnisme appelle une interprétation de la conditionalisation, mais encore qu'il en propose effectivement une (dont il soutient en outre qu'elle ne se heurte pas au paradoxe de Humphreys). En un sens, c'est bien le cas. Toutefois, nous allons montrer que cette proposition n'est pas formellement acceptable. En d'autres termes, nous allons montrer que McCurdy ne contient pas de proposition qui ait exactement la forme d'une interprétation de la conditionalisation. Nous le montrons en deux temps : d'abord une analyse de la notion d'interprétation de la conditionalisation, ensuite un examen de la proposition de Humphreys à l'aune des résultats de cette première analyse. Dans une troisième et dernière sous-section, nous proposons une interprétation propensionniste de la conditionalisation.

6.3.1 Interpréter la conditionalisation, interpréter le calcul des probabilités

Nous avons vu plus haut que les résultats de trivialité de Lewis (1976) suggèrent que conditionaliser consiste à redéfinir la fonction de probabilités elle-même – plutôt qu'à modifier l'argument pour une fonction inchangée. Si l'on suit cette suggestion, il vient l'idée selon laquelle une interprétation du calcul des probabilités dans son ensemble – c'est-à-dire en tant qu'il est étendu aux probabilités conditionnelles – se compose de :

1. une interprétation des probabilités absolues ;
2. une analyse de la façon dont la conditionalisation redéfinit une fonction de probabilités donnée.

A titre d'illustrations, nous considérons le fréquentisme et le subjectivisme – dont il n'est pas douteux qu'ils ont la forme d'interprétations du calcul des probabilités – et nous montrons qu'ils ont bien précisément cette forme que nous venons de mettre au jour. Pour ce qui est, d'abord, du fréquentisme, il consiste à :

1f. interpréter les probabilités absolues comme des fréquences relatives dans une suite[46] de réalisations d'une expérience aléatoire. Les probabilités sont alors relatives à la suite S qu'on considère ;

[46]Les hypothèses relative à la nature de cette suite (finie, infinie, réelle, hypothétique...) varient avec les différents fréquentismes.

2f. analyser la conditionalisation de la façon suivante : la conditionalisation redéfinit la fonction de probabilités relative à une suite de réalisations S, comme la fonction de probabilités relative à la suite S' qu'on obtient quand on ne garde de S que les réalisations pour lesquelles le conditionnant est satisfait.[47]

Pour ce qui est, maintenant, du subjectivisme, il consiste à

1s. interpréter les probabilités comme des degrés de croyance rationnelle. Ces degrés de croyance dépendent de l'individu I qu'on considère et du stock d'informations C dont celui-ci dispose ;

2s. considérer que la conditionalisation par une proposition P revient à ajouter P au stock d'informations dont dispose l'individu qu'on considère. Autrement dit, la conditionalisation par P revient à substituer à la fonction qui mesure les degrés de croyances de l'individu sous le stock d'informations C, la fonction qui mesure ses degrés de croyance si P est adjoint à C.

Maintenant, les exemples fréquentiste et subjectiviste donnent matière à raffiner notre analyse de la notion d'interprétation du calcul des probabilités. Les deux cas, en effet, invitent à distinguer entre deux choses :

- ce relativement à quoi une fonction de probabilités est définie : une suite d'expériences aléatoires dans le cas du fréquentisme, un individu possédant un certain ensemble de connaissances dans le cas du subjectivisme. Conformément à la terminologie introduite dans la sous-section 6.1.1 pour le propensionnisme, nous parlerons de « conditionnant fondamental ». A compter de maintenant, il sera représenté au moyen d'un indice figurant en bas à droite du symbole représentant une fonction de probabilités ;
- ce par quoi on conditionalise : un évenement générique dans le cas du fréquentisme, une proposition dans le cas du subjectivisme. Il s'agit de l'événement conditionnant, et nous le ferons figurer à la suite de la barre verticale traditionnellement utilisée pour représenter la conditionalisation.

Sous cette distinction, interpréter la conditionalisation revient précisément à indiquer comment un événement conditionnant redéfinit le conditionnant fondamental. En termes plus rigoureux, ce que nous soutenons s'énonce comme suit : la conditionalisation s'interprète comme une fonction dont le domaine est le produit cartésien de l'ensemble des conditionnants fondamentaux et

[47] Il conviendrait d'expliquer mieux ce que l'on entend par "satisfait". Mais cela nous entraînerait trop loin sur le terrain de la nature des "événements génériques" pour les bénéfices escomptés. Ceux-ci nous semblent en effet faibles dans la mesure où l'idée que nous visons est déjà bien comprise intuitivement.

de l'ensemble des événements conditionnants, et le co-domaine est un sous-ensemble de l'ensemble des conditionnants fondamentaux.

A titre d'illustrations de cette thèse, nous revenons aux interprétations fréquentiste et subjectiviste du calcul des probabilités et montrons que l'interprétation de la conditionalisation y a bien la forme que nous venons de mettre au jour :

– dans le cas du fréquentisme, si on note **S** l'ensemble des suites de réalisations d'une expérience aléatoire donnée et **E** l'ensemble des événements génériques que les membres de ces suites peuvent satisfaire, la fonction qui interprète la conditionalisation est :

$$\begin{array}{rcccc} c_f & : & \mathbf{S}\times\mathbf{E} & \longrightarrow & \mathbf{S} \\ & & (S,E) & \longmapsto & S' \end{array}$$

où S' est la suite que l'on obtient en ne retenant de S que ses éléments qui satisfont E.

– dans le cas du subjectivisme, notons **I** l'ensemble des individus, **P** l'ensemble des propositions d'un langage $\mathcal{L}$ et **C** l'ensemble des parties de **P**. La fonction qui interprète la conditionalisation est alors :

$$\begin{array}{rcccc} c_s & : & (\mathbf{I}\times\mathbf{C})\times\mathbf{P} & \longrightarrow & \mathbf{I}\times\mathbf{C} \\ & & ((I,C),P) & \longmapsto & (I,C') = (I, C\cup\{P\}). \end{array}$$

De façon plus générale, il apparaît qu'une interprétation du calcul des probabilités se compose précisément de :

1. une interprétation des probabilités absolues. Cette interprétation doit spécifier en particulier :
   (a) la nature des conditionnants fondamentaux ;
   (b) la nature des objets sur lesquels les fonctions de probabilités sont définies ;
2. une spécification de la fonction qui interprète la conditionalisation.

Armés de cette analyse, nous revenons à McCurdy (1996) et mettons au jour l'interprétation des probabilités qui y est adoptée.

6.3.2 L'interprétation adoptée par McCurdy

Ce à propos nous avons vu au début de la section 6.2 Humphreys et McCurdy s'accorder – et s'accorder avec la caractérisation proposée dans la section 5.1 – implique que le propensionnisme est d'abord :

1p. une interprétation des probabilités absolues, dans le cadre de laquelle :
   (a) les ensembles de conditions physiques d'arrière-plan jouent le rôle de conditionnants fondamentaux ;

(b) les objets sur lesquels les fonctions de probabilité sont définies sont les événements singuliers que ces ensembles de conditions d'arrière-plan sont susceptibles de produire.

Dès lors, une interprétation propensionniste du calcul des probabilités dans son ensemble doit comporter, à titre d'interprétation de la conditionalisation, la définition d'une fonction qui associe à tout couple constitué d'un ensemble de conditions d'arrière-plan et d'un événement que cet ensemble est susceptible de produire, un nouvel ensemble de conditions d'arrière-plan. Mettre au jour l'interprétation du calcul des probabilités qui sous-tend McCurdy (1996) revient donc précisément à mettre au jour comment il définit cette fonction.

Notons pour commencer que McCurdy (1996) n'aborde pas la question de l'interprétation de la conditionalisation de front et dans des termes similaires à ceux que nous venons de mettre au jour. Plutôt que d'expliciter une fonction qu'il destine à interpréter la conditionalisation, McCurdy fait fond sur le sens communément attribué à certaines expressions par les philosophes des probabilités. Ainsi évoque-t-il « la propension en $t1$ (du système satisfaisant les conditions B_{t1}) à produire un photon transmis en $t3$ *relativement au fait que* [*conditional upon*] il produit un photon qui frappe le miroir en $t2$ »[48], « la propension en $t1$ que le système produise l'événement T_{t3} *étant donné que* l'événement I_{t2} est également produit »[49], ou même « la propension que le système produise les deux événements futurs *de la manière spécifiée* »[50].

Maintenant, en nous appuyant sur les positions défendues dans McCurdy (1996), nous avons été amenés à soutenir que la relation entre un événement conditionnant et le conditionnant fondamental qu'il vient modifier n'est ni de nature physique, ni de nature épistémique. Positivement, nous l'avons caractérisée comme logique et nous l'avons décrite de la façon suivante : l'événement conditionnant vient modifier le conditionnant fondamental dans l'exacte mesure de ce que son occurrence implique relativement à la description de ce conditionnant. Selon cette lecture, conditionaliser revient à substituer à la description initiale du conditionnant fondamental, la description plus précise qui découle logiquement de la prise en compte de ce qu'il produit l'événement conditionnant.

Mais essayons d'être plus rigoureux. Formellement, la description d'un ensemble de conditions physiques peut être considérée comme une conjonction finie de propositions élémentaires. Par ailleurs, ce que la logique « prend en compte », ce sont des formules – et donc en l'occurrence ici la lettre de proposition qui représente l'énoncé selon lequel l'événement conditionnant

[48]McCurdy (1996) p. 109. C'est nous qui soulignons.
[49]McCurdy (1996) p. 109.
[50]McCurdy (1996) p. 109.

advient. Avec **P** l'ensemble des lettres de propositions et **Cj** l'ensemble des conjonctions finies de lettres de **P**, la fonction qui interprète la conditionalisation semble donc être pour McCurdy :

$$\begin{array}{lcll} c_{mc} & : & \mathbf{Cj} \times \mathbf{P} & \longrightarrow \quad \mathbf{Cj} \\ & & (p_1 \wedge p_2 \wedge \ldots \wedge p_n, p) & \longmapsto \quad p_1 \wedge \ldots \wedge p_n \wedge p. \end{array}$$

La thèse selon laquelle cette interprétation de la conditionalisation est implicite chez McCurdy est étayée par les notations qu'il utilise. Ces notations, en effet, ne donnent pas à voir de différence entre conditionnants fondamentaux et événements conditionnants. Les entités des deux types sont représentées à droite du symbole usuel pour la conditionalisation, et le symbole représentant un événement conditionnant est simplement accolé au symbole qui représente les conditions d'arrière-plan : « $Pr_{t1}(T_{t3}|I_{t2}B_{t1})$ ». En outre, McCurdy considère qu'on définit bien ainsi une nouvelle fonction de probabilités, relative à un nouveau conditionnant fondamental :

> Cette fonction est relative à l'ensemble de conditions d'arrière-plan B_{t2} qui consiste dans les conditions exprimées dans B_{t1} ainsi que la condition additionnelle que l'événement I_{t2} est advenu en $t2$ ... : $Pr_{t2}(T_{t3}|B_{t2}) = Pr_{t1}(T_{t3}|I_{t2}B_{t1})$. [51]

En définitive, la thèse selon laquelle McCurdy propose d'interpréter la conditionalisation comme la conjonction d'une proposition à la description de l'ensemble de conditions physique initial semble adéquate.

Maintenant, si la formalisation de l'interprétation du calcul des probabilités à laquelle McCurdy adhère est adéquate, elle rend patente une difficulté. Cette difficulté est la suivante : la fonction qui interprète la conditionalisation ne prend pas pour arguments les interprétations propensionnistes des conditionnants fondamentaux et des événements conditionnants, mais des *descriptions* de ces réalités. En d'autres termes, la fonction c_{mc} n'est pas définie sur le produit cartésien de l'ensemble des ensembles de conditions physiques et de l'ensemble des événements singuliers ; elle est définie sur le produit cartésien de l'ensemble des *descriptions* des ensembles de conditions physiques et de l'ensemble des propositions selon lesquelles un certain événement singulier est advenu. L'interprétation proposée pour la conditionalisation n'est donc pas, à la rigueur, compatible avec l'interprétation propensionniste des probabilités absolues telle que nous l'avons présentée au début de ce paragraphe. Dans ces conditions, la proposition de McCurdy pour l'interprétation de la conditionalisation n'est pas satisfaisante, au sens précis où sa forme n'est pas correcte. Il nous revient donc de construire une interprétation de la conditionalisation formellement correcte.

[51] McCurdy (1996) p. 112.

6.3.3 Construction d'une interprétation propensionniste de la conditionalisation

6.3.3.1 Position de la question

Nous avons proposé une formalisation de l'interprétation du calcul des probabilités qui sous-tend McCurdy (1996) et soutenu que sa forme est problématique. Plus précisément, il est apparu que la nature des objets impliqués dans l'interprétation propensionniste des probabilités absolues n'est pas compatible avec l'interprétation envisagée pour la conditionalisation. Deux voies s'offrent alors à nous pour construire une interprétation propensionniste du calcul des probabilités qui ne souffre pas de cette incohérence :

- soit modifier l'interprétation des probabilités absolues dans un sens tel que (a) les fonctions de probabilités absolues sont relatives à des *descriptions* d'ensembles de conditions physiques et (b) elles prennent pour arguments des formules représentant des énoncés selon lesquels certains événements sont advenus ;
- soit modifier l'interprétation de la conditionalisation, et proposer une fonction d'interprétation qui associe un nouvel ensemble de conditions physiques à tout couple constitué d'un ensemble de conditions physiques et d'un événement qu'il est susceptible d'engendrer.

Il nous semble clair que c'est la seconde voie qui doit être suivie si l'on veut proposer une interprétation véritablement *propensionniste* du calcul des probabilités. En effet, et ainsi que nous l'avons déjà indiqué, l'idée selon laquelle une fonction de probabilités est relative à un ensemble de conditions physiques – et donc pas à une description de cet ensemble – est essentielle au propensionnisme (et d'ailleurs reconnue comme telle tant par Humphreys et que par McCurdy). C'est donc bien une nouvelle interprétation de la conditionalisation que nous allons nous attacher à proposer.

On doit pouvoir envisager de nombreuses fonctions associant un nouvel ensemble de conditions physiques au couple constitué d'un ensemble de conditions physiques et d'un événement que cet ensemble est susceptible d'engendrer. Il convient donc de réduire l'espace des fonctions candidates, en s'appuyant sur les propriétés de la conditionalisation bayesienne. Une propriété retiendra particulièrement notre attention, et guidera une grande partie de notre analyse :

> (PC) Pour toute fonction de probabilité P et tout A du domaine de définition de P tel que $P(A) \neq 0$, $P(A|A) = 1$.

Il en découle qu'une fonction c_p n'est une interprétation propensionniste de la conditionalisation que si :

> Pour tout ensemble de conditions physiques B_{t1} et tout événement A_{t2} que B_{t1} est susceptible de produire, $Pr_{c_p(B_{t1},A_{t2})}(A_{t2}) = 1$.

Plus explicitement : la fonction de probabilités relative à l'ensemble de conditions physiques $c_{p'}(B_{t1}, A_{t2})$ associe à l'événement A_{t2}, la valeur 1.

6.3.3.2 Contre les interprétations de la conditionalisation comme un « saut temporel »

La propriété (PC) suffit à invalider les propositions d'interprétation de la conditionalisation comme un « saut temporel »[52] menant de l'ensemble de conditions physiques initial à ce qu'il est devenu à un certain instant dépendant de l'instant $t2$ qui caractérise l'événement conditionnant. En vue de dire les choses plus clairement, on peut adopter la convention suivante : un ensemble de conditions physiques B_{t1} est l'état du système physique B à l'instant $t1$. Moyennant cette convention, les interprétations de la conditionalisation comme un saut temporel en font une fonction qui à B_{t1} associe l'état du système B à un instant $t2'$ postérieur à $t1$ et dont la définition exacte dépend de l'évenement conditionnant A_{t2}.

Considérons par exemple le cas où $t2'$ est défini comme l'instant $t2^-$ qui précède immédiatement $t2$. Selon la proposition que nous envisageons, la conditionalisation sur A_{t2} redéfinit une fonction de probabilité $Pr_{B_{t1}}$ comme la probabilité $Pr_{B_{t2^-}}$ relative à ce qu'est devenu B_{t1} en $t2^-$. Or, il n'y a pas de raison de supposer que la propension qui tend à réaliser A_{t2} aura acquis une valeur de 1 par le seul fait du passage du temps entre $t1$ et $t2^-$. Dans les cas où A_{t2} n'advient pas, finalement, en $t2$, il semble même clair que la propension de B_{t2^-} à réaliser A_{t2} ne vaut pas 1.

Une solution de cette difficulté pourrait consister à ne pas définir $t2'$ comme $t2^-$, mais par exemple comme le premier instant postérieur à $t1$ et pour lequel la propension du système à réaliser A_{t2} vaut 1. Toutefois un tel instant n'existe pas toujours. Du coup, il n'est pas possible de définir de façon générale un interprétation de la conditionalisation comme saut temporel qui rende compte (PC) ; l'idée d'interpréter la conditionalisation comme un saut temporel doit donc être rejetée.

Le rejet des interprétations de la conditionalisation par un saut temporel repose, au fond, sur le fait que la propension de certains système à réaliser certains événements n'atteint jamais la valeur 1. Dans ces conditions, jouer sur le paramètre temporel de la définition des conditionnants fondamentaux

[52] L'expression est utilisée par Humphreys pour décrire les « interprétations de renormalisation » des probabilités conditionnelles (Humphreys (2004) p. 672).

ne suffit jamais à rendre compte de (PC). Positivement, il est nécessaire de considérer que *le système lui-même* varie avec la conditionalisation.

6.3.3.3 Première tentative

L'idée qui se présente immédiatement nous semble être la suivante : faire de B' le système le plus similaire à B parmi ceux pour lesquels A_{t2} advient. Ainsi, de même que la conditionalisation fréquentiste consiste à ne prendre en compte que les réalisations de l'expérience aléatoire de référence pour lesquelles le conditionnant advient, de même que la conditionalisation subjectiviste consiste à apprendre que le conditionnant advient, de même la conditionalisation propensionniste consisterait à passer à un système dans lequel le conditionnant advient. Sous cette définition de B', (PC) est bien satisfaite puisque la propension de B' qui tend à réaliser A_{t2} vaut 1.

A ce point, toutefois, une question se fait jour : certes la propension de B' qui tend à réaliser A_{t2} vaut 1 si A_{t2} advient, mais à quel instant ? Il semble clair que l'occurrence effective de A_{t2} assure une probabilité de 1 à partir de $t2$ inclus. Mais qu'en est-il des instants qui précèdent $t2$? En particulier, qu'en est-il de $t1$?

Il existe des cas pour lesquels la réponse est affirmative. Considérons par exemple un système S de transmission de photons du même type celui que Humphreys imagine, et programmé en $t0$ pour produire infailliblement un photon en $t2$. Dans ce système, un photon est produit en $t2$ – l'événement P_{t2} advient – et pour tout instant $t0 < t1 < t2$, la propension de S_{t1} à produire P_{t2} vaut bien 1.

La réponse par l'affirmative, toutefois, ne peut pas être généralisée. Imaginons en effet que : pour tout système B, tout événement A_{t2} qui advient dans B et tout instant $t1$ antérieur à $t2$, $Pr_{B_{t1}}(A_{t2}) = 1$. Alors, toutes les probabilités objectives d'événements singuliers sont triviales (égales à 0 ou à 1) et les probabilités non triviales ne représentent que de l'ignorance. En d'autres termes, l'univers considéré est déterministe – sous une définition ici évidente de ce terme. Sans même se prononcer sur la question de savoir si l'univers actuel est ou non déterministe, on peut faire valoir que le propensionnisme n'aurait pas de sens dans un tel univers.[53] Le développement d'une interprétation propensionniste du calcul des probabilités présuppose en effet qu'il existe des probabilités d'événements singuliers non triviales.

[53]Milne formule cette remarque (Milne (1986) p. 131).

6.3.3.4 Seconde tentative

On peut envisager de rendre la proposition formulée dans le dernier paragraphe recevable en la raffinant. La conditionalisation ne ferait plus varier le seul système de référence ; elle ferait également varier *l'instant* de référence. Plus précisément, l'idée serait que la conditionalisation associe à un couple $((B, t1), A_{t2})$, le couple $(B', t3)$ où B' serait le système le plus similaire à B parmi ceux dans lesquels A advient et $t3$ un instant (par exemple le premier instant) tel que $Pr_{B'_{t3}}(A_{t2}) = 1$. Notons qu'un tel instant $t3$ existe toujours puisque A_{t2} advient dans B' par définition de B'.

Il nous semble que cette nouvelle proposition se heurte à deux difficultés. En premier lieu, dans l'hypothèse indéterministe visée par le propensionnisme, elle implique qu'une probabilité conditionnelle relative à un ensemble de conditions physiques défini en $t1$ n'est pas toujours bien définie à $t1$. En effet, s'il existe (au moins) un système B' tel que la propension de B' à réaliser A_{t2} ne vaut ni 0 ni 1 en t_1, alors « le système le plus similaire à B parmi ceux dans lesquels A_{t2} advient » n'a pas de référent en $t1$. Or, il nous semble qu'on peut raisonnablement exiger d'une interprétation de la conditionalisation que l'interprétation d'une probabilité conditionnelle soit définie dès l'instant où l'est cette probabilité conditionnelle.

En second lieu, sous la proposition que nous envisageons dans le présent paragraphe, la conditionalisation fait varier *à la fois* le système et l'instant du temps qui composent l'ensemble de conditions physiques initial. Nous avons montré dans le paragraphe précédent qu'il est nécessaire que le système varie. Un souci d'économie conceptuelle impose alors de ne proposer une interprétation du type de celle que nous envisageons que s'il s'avère de s'en tenir à une variation du système *seul*. Or, précisément, la discussion qui précède permet d'envisager une proposition d'interprétation de la conditionalisation comme faisant varier le seul système de référence et qui rend compte de la propriété (PC) de la conditionalisation.

6.3.3.5 L'interprétation proposée

La proposition que nous formulons est celle de considérer que conditionaliser par A_{t2} revient à substituer au système B du conditionnant fondamental, le système B' qui lui est le plus similaire parmi ceux pour lesquels la probabilité de A_{t2} vaut 1 *à l'instant* $t1$ *de référence*. Autrement dit, et si on note $\mathbf{C}$ l'ensemble des conditionnants fondamentaux – qui, rappelons-le, sont de la forme B_{ti} – et $\mathbf{E}$ l'ensemble des événements conditionnants, la conditionalisation est interprétée par la fonction suivante :

$$\begin{array}{rcccc} c_p & : & \mathbf{C} \times \mathbf{E} & \longrightarrow & \mathbf{C} \\ & & (B_{ti}, A_{tj}) & \longmapsto & B^{\star}_{ti} \end{array}$$

où $B^\star$ est le système le plus similaire à B parmi ceux pour lesquels la probabilité de A_{tj} vaut 1 à l'instant ti.
Par construction, (PC) est satisfaite par construction de cette interprétation de la conditionalisation. En outre, l'interprétation d'une probabilité conditionnelle est bien définie à l'instant où l'est cette probabilité elle-même. Dans ces conditions, il est possible d'aller plus loin dans la discussion de cette proposition, et d'abord d'essayer de la caractériser plus précisément.

Minimalité de l'interprétation proposée. La proposition que nous formulons se caractérise d'abord par ce qu'elle est minimale à deux titres. En premier lieu, et contrairement à la proposition envisagée dans le dernier paragraphe, elle implique que la conditionalisation fait varier le seul système physique de référence, à l'exclusion de l'instant du temps auquel on le considère. Autrement dit, elle ne fait varier que ce dont on a vu qu'il était *nécessaire* qu'il varie pour qu'il soit rendu compte de (PC).

En second lieu, l'interprétation que nous proposons est telle que la différence entre le système qui compose l'ensemble de conditions physiques initial et le système qui compose l'image de l'ensemble de conditions physiques initial est minimale. En effet, le système après conditionalisation est celui qui est le plus similaire au système initial parmi ceux qui ont la propriété qui garantit que (PC) est satisfaite.

Systèmes possibles. Si la minimalité (dans les deux sens que nous venons d'expliciter) est à mettre au crédit de la proposition que nous venons de formuler, l'introduction de systèmes possibles paraît plutôt devoir être mise à son débit, surtout dans le contexte du réalisme ontologique dont le propensionnisme est souvent corrélatif.[54]. Le paragraphe qui commence vise à montrer que cette apparence est trompeuse et que la considération, à côté des systèmes actuels, de systèmes seulement possibles n'invalide pas la proposition que nous formulons.

Il convient de souligner ici d'abord que l'introduction de systèmes possibles ne peut pas être retenue contre l'interprétation que nous proposons au profit de rivales éventuelles. En effet, il est apparu dès le paragraphe 6.3.3.2 de la présente sous-section qu'il n'est pas d'interprétation de la conditionalisation qui fasse l'économie de l'introduction de nouveaux systèmes – à côté du système actuel qui déterminent les probabilités à interpréter.

Indépendamment de la prise en compte des rivales éventuelles de l'interprétation que nous proposons, il nous semble qu'il existe deux sens où l'introduction de la notion de système possible pourrait poser spécifiquement

[54]Sur ce point, voir la sous-section 5.2.1.

problème. Selon le premier, il conviendrait de s'opposer par principe à l'idée selon laquelle le calcul des probabilités engage une référence à des objets réels non observables. A cette première objection, nous répondons en faisant valoir que celui qui se pose la question de l'interprétation propensionniste de la conditionalisation a déjà accepté l'interprétation propensionniste des probabilités absolues. Il a donc déjà accepté la notion des propensions, déjà rencontré des objections similaires et déjà considéré qu'elles ne méritaient pas qu'il s'y arrête. Autrement dit, l'objection que nous envisageons ici n'est pas opposable à celui qui tente de répondre à la question de l'interprétation propensionniste de la conditionalisation.

Il y a toutefois un second sens, plus subtil, auquel l'idée de système possible pourrait être considérée comme invalidante pour la proposition d'interprétation de la conditionalisation que nous avons formulée. En ce sens, l'objection est fondée sur un calcul coût ontologique / bénéfice théorique. Elle consiste à considérer que l'introduction des systèmes possibles est un prix trop élevé pour l'interprétation propensionniste de la conditionalisation.

Cette position, si elle est possible, résiste mal à un léger élargissement de l'horizon théorique. En effet, l'analyse des énoncés contrefactuels que l'on admet communément aujourd'hui – celle qu'ont développée en particulier Stalnaker et Lewis – repose sur la considération de mondes possibles. Or, il nous semble que le prix associé à l'introduction de systèmes possibles est négligeable une fois qu'on a accepté des mondes possibles. Dès lors, pour celui qui accepte les mondes possibles, l'introduction de systèmes possibles ne peut pas être un prix trop élevé pour disposer d'une interprétation propensionniste des probabilités conditionnelles. Au sens où nous les avons définis (dans la sous-section 4.1.1), les systèmes ne sont rien d'autre, après tout, que des portions de mondes. Réciproquement, refuser les systèmes possibles implique maintenant de refuser avec eux les mondes possibles, et donc de renoncer à l'analyse des contrefactuels. Comme les contrefactuels semblent jouer un rôle central en particulier dans les jugements causaux (juger que A a causé B, c'est considérer que B ne serait pas advenu si B n'était pas advenu), renoncer à l'analyse des contrefactuels apparaîtra précisément comme un prix trop élevé à payer pour ne pas avoir à faire référence à des systèmes non actuels. En résumé, le recours à la notion de système possible n'est pas un prix trop élevé pour résoudre le problème de l'interprétation propensionniste des probabilités conditionnelles dès lors qu'on ne considère pas ce problème isolément, mais comme appartenant à une classe de problèmes dont l'analyse des contrefactuels fait partie.

En plus de considérer des systèmes possibles, notre interprétation propensionniste de la conditionalisation suppose qu'il existe une fonction qui à chaque système associe celui qui lui est le plus similaire parmi l'ensemble des

systèmes physiques possibles. De même que les systèmes possible, la fonction qui associe à tout système celui qui lui est le plus similaire a un analogue dans le cadre de l'analyse communément admise des contrefactuels. Du coup, un raisonnement du même type que celui que nous venons de mener peut être opposé à tout rejet de notre proposition d'interprétation de la conditionalisation qui serait fondé sur une critique de l'hypothèse selon laquelle il existe une telle fonction qui à chaque système possible associe celui qui lui est le plus similaire – une « fonction de sélection ».

Fonction de sélection. Il reste toutefois deux questions que soulève l'hypothèse de l'existence d'une telle fonction de sélection. La première est celle de sa définition – et donc, au-delà, de la définition de la similarité entre systèmes elle-même. Sur ce point, il apparaît d'emblée à la fois qu'il est impossible de définir complètement cette fonction et que, dans le contexte qui nous intéresse, sa définition doit prendre en compte de manière centrale des caractéristiques propensionnistes des systèmes physiques considérés.

La seconde question soulevée par l'hypothèse d'une fonction de sélection est la suivante : pourquoi supposer que la relation de plus grande similarité entre les systèmes existants est fonctionnelle ? Autrement dit : pourquoi supposer que, pour tout système physique S, il existe un *et un seul* système le plus similaire à S parmi les systèmes possibles ? Cette question est exactement similaire à celle de Lewis à Stalnaker dans le contexte de l'analyse des conditionnels contrefactuels, et nous considérons qu'elle est pertinente. Nous reconnaissons qu'il n'y a en effet aucune raison de penser que la relation de plus grande similarité entre systèmes est fonctionnelle. Mais, d'un autre côté, nous remarquons qu'abandonner l'hypothèse d'unicité – puisque c'est le nom que Lewis lui donne – rendrait l'interprétation que nous proposons et sa discussion plus complexes encore qu'elles ne le sont déjà. A cela, nous préférons aller du plus simple au plus complexe, c'est-à-dire développer et discuter jusqu'au bout la proposition que nous avons formulée, et n'envisager de lever l'hypothèse d'unicité qu'éventuellement – s'il s'avère qu'une interprétation du type de celle que nous formulons échappe au paradoxe de Humphreys – et dans un travail ultérieur. Pour l'heure, la question qui se pose est clairement celle de savoir si l'interprétation de la conditionalisation constitue en effet une solution du paradoxe de Humphreys. C'est que nous allons essayer de déterminer maintenant.

6.4 Discussion de la proposition : Le paradoxe de Humphreys est-il résolu ?

L'interprétation de la conditionalisation que nous proposons à été construite de telle sorte qu'elle a la forme d'une interprétation propensionniste de la conditionalisation (contrairement à l'interprétation adoptée par McCurdy) et qu'elle satisfait la propriété (PC). Cela, toutefois, laisse ouverte la question de savoir si nous avons résolu le paradoxe de Humphreys. Posée en d'autres termes, cette question est celle de savoir si la proposition formulée dans la dernière section permet de rendre compte des propriétés de la conditionalisation bayésienne.

Dans la présente section, nous commençons par aborder cette question de manière négative et sous la forme suivante : l'interprétation que nous proposons n'est-elle pas une analyse des probabilités de conditionnels – plutôt qu'une interprétation des probabilités conditionnelles ? Cette stratégie a deux motivations. D'une part, nous verrons qu'il y a effectivement de bonnes raisons de penser que nous avons construit une interprétation de probabilités de conditionnels plutôt qu'une interprétation des probabilités conditionnelles. D'autre part, si nous montrions que nous avons effectivement produit une interprétation de probabilités de conditionnels, la question qui nous occupe serait réglée. Il serait alors inutile de traiter la question telle que se pose positivement – et dont le lecteur aura pressenti qu'il est difficile de lui apporter une réponse.

Nous montrerons toutefois que les bonnes raisons de penser que nous avons produit une interprétation de probabilités de conditionnels ne résiste pas à un examen attentif. Dans ces conditions, nous devrons nous colleter dans une seconde sous-section avec la question positive de savoir si notre interprétation est admissible.

6.4.1 Probabilités de conditionnels ?

La question de savoir si les propriétés capturées ne sont pas celles des probabilités de conditionnels, plutôt que celles des probabilités conditionnelles, est une question pertinente quelle que soit la proposition d'interprétation des probabilités conditionnelles que l'on considère. On sait en effet depuis Lewis (1976) que les probabilités conditionnelles et les probabilités de conditionnels ne coïncident que dans certains cas triviaux.[55] Mais, ainsi que nous l'avons annoncé, la question se pose de manière particulièrement aiguë dans le cas qui nous occupe, puisqu'il semble exister de bonnes raisons de penser que

[55]Lewis (1976) p. 300–303.

notre proposition a pour objets des probabilités de conditionnels de Stalnaker, plutôt que des probabilités conditionnelles. Nous présentons ces raisons maintenant.

6.4.1.1 En faveur de l'idée selon laquelle nous aurions produit une analyse de probabilités de conditionnels

L'argument de la ressemblance. On peut commencer ce paragraphe en dressant un parallèle entre notre proposition d'interprétation de la conditionalisation et l'analyse des conditionnels contrefactuels par Stalnaker. D'un côté, notre analyse

1. repose sur :
   (a) l'hypothèse selon laquelle il existe des systèmes possibles non actuels ;
   (b) l'hypothèse selon laquelle il existe une fonction de sélection qui à chaque système S et chaque propriété P associe le système le plus similaire de S parmi ceux qui ont la propriété P.
2. a pour objet les probabilités dans le système le plus similaire au système initialement considéré, parmi ceux pour lesquelles une certaine propriété – la propriété (PC) – est satisfaite.

De l'autre côté, l'analyse de Stalnaker :

1. repose sur :
   (a) la référence au concept de monde possible et l'utilisation de « l'appareil théorique tout prêt [qui lui correspond et] sur lequel construire une une théorie sémantique formelle »[56] ;
   (b) l'hypothèse selon laquelle il existe « une *fonction de sélection*, *f*, qui prend une proposition et un monde possible comme arguments et un monde possible comme valeur »[57] et qui est telle que la proposition antécédente est vraie dans le monde sélectionné et que celui-ci par ailleurs « *diffère minimalement* du monde actuel »[58].
2. consiste à considérer que le conditionnel $A > B$ est vrai dans le monde α si et seulement la proposition B est vraie dans le monde $f(A, \alpha)$[59] – c'est-à-dire dans le monde qui diffère le moins de α parmi ceux dans lesquels la proposition A est vraie. Dès lors, la probabilité d'un conditionnel $A > B$ semble être la probabilité de la proposition B dans le

[56] Stalnaker (1968) p. 103.
[57] Stalnaker (1968) p. 103.
[58] Stalnaker (1968) p. 104.
[59] Voir Stalnaker (1968) p. 103.

monde le plus similaire au monde actuel parmi ceux qui ont la propriété que la proposition A y est vraie.

Cette présentation en parallèle fait apparaître que, sous l'interprétation que nous proposons, la probabilité conditionnelle $p(B|A)$ ressemble fortement à la probabilité du conditionnel $[(p(A) = 1) > B]$ sous l'analyse que Stalnaker donne de ce conditionnel. Deux différences doivent toutefois être notées : d'une part, nous nous intéressons à des probabilités d'événements là où Stalnaker considère des propositions ; d'autre part, nous parlons d'ensembles de conditions physiques là où Stalnaker parle de mondes. Il semble toutefois qu'aucune de ces deux différences ne suffit à invalider le rapprochement de l'interprétation que nous proposons et des probabilités de conditionnels de Stalnaker, mais que toutes deux invitent à préciser les termes de ce rapprochement.

Concernant d'abord la distinction entre événements et propositions, la difficulté disparaît si l'on considère que la probabilité d'un événement singulier comme la probabilité de la proposition que cet événement advient. La proposition que nous formulons ici appelle deux remarques. En premier lieu, il convient de souligner que nous ne proposons pas ici un principe de correspondance entre probabilités objectives et probabilités subjectives. Nous nous contentons de remarquer qu'il est possible de considérer qu'une fonction de probabilités *objective* prend pour arguments des propositions plutôt que des événements – le rapport entre la probabilité d'un événement et celle de la proposition qu'il advient étant d'égalité. En second lieu, on notera que proposer de considérer que les arguments d'une fonction de probabilités objectives sont des propositions ne nous expose pas au reproche que nous adressions à McCurdy dans la sous-section 6.3.2. En effet, ce que nous reprochions alors à McCurdy était de ne pas interpréter la conditionalisation autrement que comme une conjonction de propositions. De notre côté, nous avons proposé une interprétation de la conditionalisation en termes physiques, et ne revenons à une formulation en termes de propositions dans le cadre de l'interprétation proposée – et seulement pour des raisons de commodité.

Sous la convention que la probabilité objective d'un événement singulier peut être considérée comme la probabilité objective de la proposition que cet événement advient, l'interprétation que nous proposons semble conduire à considérer la probabilité conditionnelle $Pr(B|A)$ comme la probabilité du conditionnel dont

– l'antécédent est la proposition selon laquelle la probabilité de la proposition que A advient vaut 1 – soit $Pr($« A advient »$) = 1$;

– le conséquent est la proposition que B advient – soit « B advient »,

soit la probabilité : $Pr[(Pr($ « A advient »$) = 1) >$« B advient »$]$.

Concernant maintenant la distinction entre ensembles de conditions physiques et mondes, elle n'est en fait qu'une variante de la distinction précédente : tandis qu'un ensemble de conditions physiques est constitué d'instanciations de propriétés par des individus, un monde possible est un ensemble de propositions. Dans ces conditions, on traite la difficulté d'une façon similaire à celle dont on a levé la difficulté précédente. Plus explicitement, on conviendra qu'il est possible de considérer que le conditionnant fondamental d'une fonction de probabilités est la description d'un ensemble de conditions physiques – plutôt que cet ensemble lui-même. Il apparaît alors que, pas plus que la distinction entre probabilités d'événements et probabilités de propositions, la distinction entre ensembles de conditions physiques et mondes possibles n'empêche de rapprocher les probabilités conditionnelles telles que nous les analysons des probabilités de conditionnels de Stalnaker.

L'argument de l'*imaging*. Il existe un second argument en faveur de l'idée selon laquelle notre interprétation, si elle vise les probabilités conditionnelles, atteint des probabilités de conditionnels. Ces arguments reposent sur le résultat établi dans la section « Probabilities of Stalnaker Conditionals » de Lewis (1976). Selon ce résultat, la probabilité $p(A > B)$ d'un conditionnel de Stalnaker $A > B$ est la probabilité de B pour la fonction de probabilités p^i_A qui résulte de la révision de p par *imaging* sur A.

L'*imaging* est une méthode de révision des fonctions de probabilités rivale de la conditionalisation bayésienne.[60] De même que la conditionalisation bayésienne, elle est telle que ce relativement à quoi on révise la fonction de probabilités initiale – ce que nous avons appelé « conditionnant » dans le cas de la conditionalisation bayésienne – a une probabilité de 1 pour la fonction de probabilités révisée.[61] Autrement dit, l'*imaging* satisfait la même propriété (PC) que la conditionalisation bayésienne. Il en découle que notre interprétation de la conditionalisation, satisfaisant (PC), pourrait être une interprétation de l'*imaging* plutôt que de la conditionalisation bayésienne.

Par ailleurs, de même que la conditionalisation bayésienne, la conditionalisation *imaging* « révise la fonction de probabilités données autant qu'il est nécessaire pour rendre la proposition [sur laquelle on conditionnalise] certaine, mais pas plus »[62]. Les deux méthodes de révision sont donc minimales, mais en deux sens différents :

[60]Pour une définition, voir Lewis (1976) p. 310. Nous ne présentons dans ce paragraphe que les caractéristiques de l'*imaging* qui sont pertinentes relativement à la discussion que nous y menons.

[61]Cf. Lewis (1976) p. 313.

[62]Lewis (1976) p. 311.

> Prendre l'*imaging* de p sur A donne une révision minimale en ce sens : contrairement à toutes les autres révisions de p qui rendent A certain, cela n'implique aucun mouvement gratuit de probabilités d'un monde vers un monde différent. Conditionaliser p par A donne une révision minimale en ce sens différent : contrairement à toutes les autres révisions de p qui rendent A certain, cela ne distort pas le profile des quotients, égalités et inégalités de probabilités parmi les propositions qui impliquent A. [63]

Or Bernard Walliser et Denis Zwirn montrent dans Walliser et Zwirn (2002) que la distinction entre ces deux sens de minimalité est la contre-partie probabiliste de la distinction établie dans Katsuno et Mendelzon (1992) entre deux types de changements de croyances : la mise à jour [*update*] d'une part et la révision [*revision*] d'autre part.[64] La mise à jour doit être adoptée dans les cas où l'agent considéré apprend qu'un changement a eu lieu dans le monde, la révision dans les cas où il apprend une nouvelle information sur un monde inchangé :

> La *mise à jour* consiste à rendre actuelle la base de connaissances quand le monde qu'elle décrit change. [...] La *révision* est utilisée quand nous obtenons une nouvelle information sur un monde statique.[65]

Dans ces conditions, l'hypothèse selon laquelle notre proposition d'interprétation capturerait des cas d'*imaging* plutôt que des instances de conditionalisation bayésienne gagne encore en plausibilité. En effet, rappelons que le propensionnisme est une théorie selon laquelle les probabilités sont des caractéristiques objectives du monde physique. Dès lors, il semble raisonnable de penser qu'il conduit à interpréter le changement des fonctions de probabilités comme résultant de changements dans le monde. Surtout, c'est bien le cas pour l'interprétation propensionniste de la conditionalisation que nous envisagions plus haut. Ce changement correspond donc à une mise à jour plutôt qu'à une révision, et corrélativement relève de l'*imaging* plutôt que de la conditionalisation bayésienne.

[63] Lewis (1976) p. 311.

[64] Les résultats obtenus par Walliser et Zwirn sont en fait à la fois plus nombreux et plus nuancés. En effet, ils considèrent plusieurs systèmes d'axiomes de changement des croyances probabilistes, dont certains relèvent de la mise à jour et les autres de la révision. Pour chacun de ces systèmes, ils établissent qu'il est adéquatement représenté par une méthode de changement des fonctions de probabilités. Pour les systèmes de mise à jour des croyances, cette méthode est une forme d'*imaging* ; pour les systèmes de révision des croyances, cette méthode est une forme de conditionalisation bayésienne.

[65] Katsuno et Mendelzon (1992) p. 183.

En définitive, l'idée selon laquelle la proposition que nous avons formulée relève de l'analyse des probabilités de conditionnels plutôt que de l'interprétation des probabilités conditionnelles est soutenue par deux arguments. D'abord, l'idée de considérer les probabilités dans l'ensemble de conditions physiques le plus similaire à l'ensemble considéré parmi ceux pour lesquels une certaine propriété est satisfaite entretient avec l'idée qui préside à l'analyse des énoncés conditionnels proposée par Stalnaker une ressemblance forte et qui résiste à une analyse de leurs différences les plus visibles. Ensuite, le fait que la probabilité d'un conditionnel est la probabilité de son conséquent après un changement de la fonction de probabilités qui rend compte de changements *dans* le monde rend plausible l'idée selon laquelle le propensionnisme nous aurait conduits à une analyse des probabilités de conditionnels plutôt qu'à une interprétation de la conditionalisation bayésienne. Avant de discuter ces arguments, il convient de s'arrêter un instant sur ce que seraient les conséquences de leur caractère concluant.

Ce qu'impliquerait que les arguments précédents soient concluants. Nous avons rappelé au début de la présente sous-section que les probabilités de conditionnels ne sont pas des probabilités conditionnelles. Nous en avons inféré que notre interprétation de la conditionalisation ne rend pas compte des propriétés de la conditionalisation bayésienne si elle est effectivement de probabilités de conditionnels. Cette inférence, toutefois, ne va peut-être pas de soi.

En effet, nous avons montré qu'il y a de bonnes raisons de penser que l'interprétation que nous proposons pour $Pr(B|A)$ en fait la probabilité du conditionnel ($Pr($ « A advient ») $= 1) >$« B advient »). Mais, d'un autre côté, les résultats de trivialité de Lewis porte précisément sur l'égalité

$$\text{(E)} \qquad p(A > B) = p(B|A).$$

Dans ces conditions, que nous proposions d'interpréter $Pr(B|A)$ comme la probabilité $Pr[(Pr($ « A advient ») $= 1) >$« B advient »] n'implique peut-être pas que nous échouons à interpréter la conditionalisation bayésienne. En d'autres termes, il se pourrait que notre proposition d'interprétation de la conditionalisation bayésienne soit admissible alors même que les deux arguments que nous venons de faire valoir seraient concluants.

L'hypothèse que nous envisageons ici, toutefois, ne résiste pas à l'examen des preuves des résultats de trivialité établis dans Lewis (1976). Plus précisément, cet examen fait apparaître que les preuves ne sont pas sensibles à la forme exacte des antécédents des conditionnels considérés. Du coup, on les reconstruit aisément pour le cas où ils sont des propositions selon lesquelles certaines probabilités valent 1. On obtient alors des absurdités analogues de

celles que Lewis met en évidence. Notre proposition doit bien être rejetée s'il est vrai qu'elle consiste à interpréter les propriétés conditionnelles comme des probabilités de conditionnels de la forme $Pr[(Pr($ « A advient ») $= 1) >$« B advient »]. Les arguments que nous venons de présenter en faveur de cette hypothèse semblent forts, mais c'est eux qu'il convient de discuter maintenant.

6.4.1.2 Contre l'idée selon laquelle nous aurions produit une analyse de probabilités de conditionnels

Contre l'argument de l'*imaging*. Nous revenons ici sur le résultat établi par Lewis dans la section « Probabilities of Stalnaker Conditionals » de Lewis (1976). Selon ce résultat, nous l'avons déjà dit, les conditionnels de Stalnaker ont des probabilités égales aux probabilités de leur conséquent pour une fonction modifiée par *imaging* sur l'antécédent. La validité du second argument avancé dans le paragraphe précédent repose tout entière sur ce résultat. Or, un examen attentif révèle qu'il n'est pas du tout évident que ce résultat soit disponible dans un cadre propensionniste.

Revenons en effet à la preuve que Lewis en donne.[66] On peut considérer qu'elle est composée de trois parties : 1) une première égalité permet de passer de la probabilité d'une proposition à des *probabilités de mondes* ; 2) les égalités centrales correspondent à un travail sur des expressions faisant intervenir des probabilités de mondes ; 3) une dernière égalité marque le retour des probabilités de mondes à la probabilité d'une proposition. C'est au niveau de 2) que se joue ce qui rend le résultat vrai. Pour que la preuve soit possible, il est donc indispensable de pouvoir travailler à ce niveau – celui des probabilités de mondes – et que les probabilités de propositions et les probabilités de mondes soient dans le même rapport que chez Lewis : la probabilité d'une proposition est la somme des probabilités des mondes dans lesquels elle est vraie.

La contre-partie propensionniste de cette propriété serait que la probabilité d'un événement soit la somme des probabilités des ensembles de conditions physiques possibles dans lesquels il advient. Or cette contre-partie n'est pas seulement fausse : elle n'a pas de sens. Dans un contexte propensionniste, en effet, les ensembles de conditions physiques *déterminent* des probabilités, ils n'*ont* pas de probabilités. S'ils en avaient une, ce serait relativement à un conditionnant plus fondamental encore et qui serait susceptible de les engendrer. Or, on voit mal ce que pourrait être un tel conditionnant. Surtout, même en acceptant l'idée d'un tel conditionnant et en considèrant la proba-

[66] Lewis (1976) p. 311.

bilité d'un événement singulier relativement à ce nouveau conditionnant, il n'y a aucune raison de penser que cette probabilité est la somme des probabilités des ensembles de conditions physiques susceptibles d'engendrer cet événement.

Si Lewis peut mobiliser l'égalité entre la probabilité d'une proposition et la somme des probabilités des mondes dans lesquels cette proposition est vraie, il semble bien que c'est en raison du caractère subjectif des probabilités qu'il considère. Dans un tel contexte, en effet, il n'y a pas d'autre sens possible à la notion de probabilité d'une proposition que celui de probabilité que le monde actuel soit tel que cette proposition y est vraie. Ce qu'on peut dire des probabilités n'est donc pas indifférent, finalement, à l'interprétation qu'on en donne. En effet, et quoi qu'il en soit des raisons exactes de la validité de la propriété mobilisée par Lewis, il est apparu que prouver un résultat analogue du sien est impossible dans un contexte propensionniste. Du coup, les arguments faisant référence aux propriétés de l'*imaging* qui ont été présentés dans le paragraphe précédent ne sont pas pertinents.

Nous venons de montrer que notre second argument en faveur de l'idée selon laquelle nous aurions proposé une analyse de probabilités de conditionnels de Stalnaker plutôt qu'une interprétation des probabilités conditionnelles devait être abandonné. Logiquement, cela ne retire rien de la pertinence des arguments de la première famille, qui consistait à faire valoir que l'interprétation que nous proposons pour $Pr(A|B)$ se présente comme la probabilité du conditionnel ($Pr($« A advient »$) = 1) >$« B advient ». Il se pourrait que cette analyse reste correcte, alors même que les propriétés des probabilités de conditionnels ne sont pas les mêmes selon que ces probabilités sont considérées comme des probabilités objectives ou comme des probabilités subjectives. Il nous semble que ce n'est pas le cas, et nous nous apprêtons à expliquer pourquoi.

Contre l'argument de la ressemblance. La critique de l'argument de l'*imaging* ouvre la voie d'une critique de l'argument de la ressemblance. En vue de le comprendre, il convient de souligner que la critique de l'argument de l'*imaging* repose finalement sur le fait que, dans le propensionnisme, il n'existe pas de rapport entre ce que pourraient être les probabilités d'ensembles de conditions physiques d'une part, et d'autre part les probabilités d'événements relatives à ces ensembles de conditions physiques.

Revenons, maintenant, à l'argument de la ressemblance. Il consiste à faire valoir que, sous l'interprétation que nous proposons, $Pr(B|A)$ s'interprète comme quelque chose qui ressemble fortement à la probabilité du conditionnel « (Pr(A)$= 1) >$B ». Maintenant, étant données les conditions de vérité des

conditionnels de Stalnaker, la probabilité de ce conditionnel (Pr(A)= 1) >B est celle de la proposition « B [est vraie] dans le (Pr(A) = 1)-monde[67] le plus similaire au monde actuel ». Cette probabilité se présente comme ce par quoi nous proposons d'interpréter $Pr(B|A)$ parce que la probabilité de la proposition « B dans le $(Pr(A) = 1)$-monde le plus similaire au monde actuel » est identique à la probabilité de la proposition B dans le $(Pr(A) = 1)$-monde le plus similaire au monde actuel.

Cette identité n'est pas problématique dans le cadre d'analyse subjectiviste. Elle le devient, en revanche, dans un cadre d'analyse propensionniste. Dans ce cadre, en effet, la probabilité de la proposition « B dans le $(Pr(A) = 1)$-monde le plus similaire au monde actuel » est relative à un conditionnant fondamental susceptible d'engendrer les ensembles de conditions physiques que décrivent les mondes possibles. Ces ensembles de conditions physiques apparaissent du côté des arguments de la fonction de probabilités. A l'inverse, la probabilité de la proposition B dans le $(Pr(A) = 1)$-monde le plus similaire au monde actuel est clairement une probabilité relative à un ensemble de conditions physiques du même type que ceux que la fonction de probabilités précédentes compte au nombre de ces arguments. Dans ces conditions, la probabilité de la proposition « B dans le $(Pr(A) = 1)$-monde le plus similaire au monde actuel » et la probabilité de B dans le $(Pr(A) = 1)$-monde le plus similaire au monde actuel ne sont plus identiques.

Finalement, nous avons fait apparaître que l'identité sur laquelle repose l'argument de la ressemblance ne vaut pas dans un cadre propensionniste. Il en découle que cet argument, qui nous a conduits à envisager que nous proposions d'interpréter les probabilités conditionnelles comme des probabilités de certains conditionnels, n'est pas concluant. Nous n'avons donc plus de bonnes raisons de considérer que nous avons proposé une analyse de probabilités de conditionnels plutôt qu'une interprétation des probabilités conditionnelles. La question qui se pose alors est celle de savoir si, positivement, l'interprétation que nous proposons est admissible – et, de manière équivalente, résout le paradoxe de Humphreys.

6.4.2 Probabilités conditionnelles ?

A la question de savoir si l'interprétation que nous proposons est admissible, nous n'avons de réponse définitive. Toutefois, nous avons deux types d'éléments de réponse à faire valoir, que nous exposons successivement. Dans un premier temps, nous revenons sur le paradoxe de Humphreys ;

[67] De façon générale, un P-monde est un monde dans lequel la proposition P est vraie.

dans le second temps, nous explorons l'analogie entre la question que nous posons maintenant et la question de savoir si le propensionnisme est une interprétation des probabilités absolues.

6.4.2.1 Retour au paradoxe de Humphreys

Nous avons montré dans le paragraphe 6.2.2.2 que le paradoxe de Humphreys est attaché à une conception des probabilités conditionnelles dont Humphreys considère qu'elle est contenue analytiquement dans la théorie propensionniste des probabilités absolues. De la même façon, les autres versions du paradoxe formel qui sont présentées dans Humphreys (2004) atteignent des conceptions de ce que la théorie propensionniste des probabilités absolues implique relativement aux probabilités conditionnelles. Contre ces théories (d'ailleurs concurrentes) des propensions conditionnelles, nous avons tenté de résoudre le paradoxe de Humphreys en construisant une *interprétation* propensionniste de la conditionalisation.

Pour que cette stratégie nous ait effectivement permis de résoudre le paradoxe de Humphreys, une condition nécessaire est que l'interprétation proposée pour la conditionalisation ne conduise à aucun des trois principes d'évaluation des probabilités conditionnelles inverses contre lesquels Humphreys a un argument formel. En vue de déterminer ce qu'il en est, on peut revenir au dispositif B imaginé par Humphreys et à la probabilité $Pr_{B_{t1}}(I_{t2}|T_{t3})$ qu'un photon émis en $t0$ frappe le miroir en $t2$ étant donné qu'il traverse le miroir en $t3$.

Sous l'interprétation que nous proposons pour la conditionalisation, cette probabilité mesure en $t1$ la propension qui tend à réaliser I_{t2} dans le système le plus similaire à B parmi ceux relativement auxquels $Pr_{t1}(T_{t3}) = 1$. Or, il nous semble que le système le plus similaire au dispositif B décrit par Humphreys parmi ceux qui donnent en $t1$ une probabilité de 1 à T_{t3} est tel que 1) les photons sont contraints physiquement de frapper le miroir et 2) le miroir est transparent. Si cette analyse est correcte, alors $Pr_{B,t1}(I_{t2}|T_{t3})$ vaut 1. L'évaluation que nous proposons est donc la même que celle que propose McCurdy. Surtout, elle n'est conforme ni au principe (CI) de Humphreys, ni au principe (ZI). Elle est conforme au principe (FP) qui veut que toutes les probabilités conditionnelles inverses valent 0 ou 1, mais il n'en découle pas que notre interprétation est soumise au paradoxe de Humphreys. En effet, (FP) ne se heurte au paradoxe de Humphreys qu'en tant qu'il implique que certaines probabilités conditionnelles inverses sont nulles.[68]

Au-delà, maintenant, de $Pr_{B_{t1}}(I_{t2}|T_{t3})$, l'interprétation que nous propo-

[68] Humphreys (2004) p. 671.

sons ne conduit pas à un principe général pour l'évaluation des probabilités conditionnelles inverses. Plus précisément, il faut toujours revenir au conditionnant fondamental pour déterminer la valeur des probabilités conditionnelles inverses et, corrélativement, celles-ci peuvent prendre leurs valeurs dans l'intervalle [0 ; 1] entier. Dans ces conditions, il n'est pas possible d'opposer à l'interprétation que nous proposons une nouvelle version formelle du paradoxe de Humphreys. En effet, une telle version formelle du paradoxe n'existe jamais qu'en référence à un principe général pour l'évaluation des probabilités conditionnelles inverses. L'absence d'un tel principe constitue donc un atout de l'interprétation que nous proposons. Elle n'implique pas, toutefois, que l'interprétation proposée est bien admissible.

6.4.2.2 Analogie avec le cas des probabilités absolues

La question de savoir si l'interprétation que nous proposons rend compte des propriétés de la conditionalisation est indiscutablement analogue de celle, traitée dans la sous-section 5.1.3, de savoir si le propensionnisme satisfait les axiomes de Kolmogorov pour les probabilités absolues. En particulier, elle se heurte à la même difficulté fondamentale que constitue l'absence d'une procédure permettant d'évaluer les probabilités – ou, pour le dire dans les termes de Gillies, le caractère non opérationaliste de la définition des probabilités sous l'interprétation proposée. On notera que cette difficulté est comme redoublée à l'occasion du passage aux probabilités conditionnelles : sous l'interprétation que nous proposons, évaluer les probabilités conditionnelles requiert non seulement de mesurer des propensions, mais encore de les mesurer pour un système seulement possible. Ce redoublement, toutefois, ne rend pas la situation plus critique : la seule distinction pertinente ici est entre l'existence et l'absence d'une procédure permettant d'évaluer les probabilités ; elle n'est pas entre moins et plus d'éléments expliquant cette absence.

Maintenant, du côté des probabilités absolues, la discussion a mis au jour un (et un seul) argument en faveur de la thèse selon laquelle le propensionnisme serait une interprétation admissible du calcul des probabilités absolues. Cet argument est celui de Lewis ; il consiste à faire valoir que les propensions, étant avec les degrés de croyance rationnelle dans le rapport énoncé par le Principe Principal, ne peuvent pas manquer de satisfaire les axiomes 1. à 3. de Kolmogorov[69]. Le Principe Principal, rappelons-le, est le principe selon lequel les degrés de croyance s'alignent sur les probabilités objectives d'événements singuliers quand celles-ci sont connues. Autrement

[69]Pour les axiomes de Kolmogorov, voir le paragraphe 5.1.3.1.

dit, le degré de croyance en A relativement à l'information que la probabilité objective de A vaut x est x.

Le Principe Principal est un principe qui fixe la valeur de probabilités subjectives conditionnelles dont le conditionné est une proposition et le conditionnant un énoncé de probabilité objective. Dès lors, produire dans le cas conditionnel un argument analogue de celui de Lewis pour le cas absolu requiert de considérer des probabilités dont le conditionné est une probabilité subjective conditionnelle et le conditionnant un énoncé de probabilité (conditionnelle) objective. Or il nous semble clair que cela n'a pas de sens. Plus précisément, il nous semble clair que n'a pas de sens la notion de probabilité deux fois conditionnelle. Dans ces conditions, l'argument de Lewis pour le cas absolu n'a pas d'équivalent dans le cas conditionnel. Pour le dire autrement : l'analogie entre la question qui nous occupe dans la présente sous-section et la question de l'admissibilité de la théorie propensionniste des probabilités absolues trouve ici une limite.

Il ne nous reste plus, alors, qu'à *supposer* que l'interprétation que nous proposons rend bien compte des propriétés de la conditionalisation bayésienne. C'est ce que nous avons fait, finalement, pour la théorie propensionniste des probabilités absolues. Surtout, il nous semble que c'est ce que fait la plupart des tenants d'une interprétation propensionniste des probabilités singulières dans le cas absolu. L'analogie entre la question posée dans la présente sous-section et la question de savoir si le propensionnisme de cas singuliers est une interprétation des probabilités absolues prend alors tout son intérêt argumentatif. Si, en effet, le propensionnisme est une interprétation admissible des probabilités absolues *par hypothèse*, alors il devient pensable d'émettre l'*hypothèse* selon laquelle l'interprétation que nous proposons pour les probabilités conditionnelles est admissible.

Assurément, notre défense de la thèse selon laquelle ce que nous avons proposé est bien une interprétation des probabilités conditionnelles est plutôt faible. Deux points, toutefois, viennent contre-balancer cette faiblesse. Le premier consiste à faire valoir que les problèmes que nous avons rencontrés dans la sous-section qui s'achève ne sont pas spécifiques de notre proposition d'interprétation propensionniste de la conditionalisation. Quelle que soit l'interprétation propensionniste de la conditionalisation qu'on adopte, on se heurte d'abord à l'impossibilité de mesures les propensions, et ensuite à ce qu'il n'existe pas d'analogue de l'argument de Lewis pour les probabilités conditionnelles. Pour ce qui concerne plus précisément l'interprétation que nous proposons, nous avons au moins montré qu'il n'y a pas de raisons de penser qu'il s'agit d'une interprétation de l'*imaging* plutôt que de la conditionalisation bayésienne.

En second lieu, nous souhaitons suggérer que l'apport du présent chapitre au débat relatif à l'interprétation propensionniste des probabilités conditionnelles réside peut-être moins dans le contenu de l'interprétation proposée que dans la façon de l'aborder. Ainsi, nous avons montré qu'une interprétation propensionniste des probabilités conditionnelles reste à formuler puis analysé ce que veut dire interpréter la conditionalisation. A la suite de cette analyse, nous avons pu approcher la tâche de formulation d'une interprétation propensionniste de la conditionalisation sur un mode qu'on pourrait qualifier de constructiviste. Il nous semble que tout cela continue de contribuer au débat sur l'interprétation propensionniste des probabilités conditionnelles quand bien même l'interprétation que nous proposons ne pourrait pas, finalement, être acceptée. Armés de la distinction entre la façon d'approcher le problème de l'interprétation propensionniste de la conditionalisation d'une part et d'autre part la solution effectivement proposée, nous revenons pour une dernière section sur la question générale du rapport entre propensionnisme et causalité, et sur l'idée plus particulière d'une théorie probabiliste de la causalité actuelle.

6.5 Propensionnisme et causalité. Deux conclusions

Dans la section qui commence, la distinction entre l'approche du problème de l'interprétation propensionniste des probabilités conditionnelles et la solution que nous proposons est prise en compte selon la modalité suivante : les conclusions sont tirées en deux temps, dont le premier dépend de l'inteprétation proposée pour la conditionalisation et le second ne dépend que de la façon dont le problème a été approché. Ainsi, le lecteur qui rejette notre proposition d'interprétation propensionniste de la conditionalisation peut concentrer son attention sur la seule seconde sous-section.

6.5.1 Probabilités conditionnelles et causalité

Si la question de la causalité – et plus précisément de son rapport avec les probabilités – n'est abordée qu'à ce point du présent chapitre, c'est que la question du rapport entre causalité et probabilités est d'abord la question du rapport entre causalité et probabilités *conditionnelles*. Aussi, aborder la question du rapport entre la causalité singulière et les probabilités objectives impliquait de traiter le problème de l'interprétation propensionniste de la conditionalisation. Dans ces conditions, la question qui se pose maintenant est la suivante : quel est le rapport entre la causalité et les probabilités

conditionnelles sous l'interprétation de la conditionalisation que nous venons de proposer ? Cette question est celle de la possibilité de proposer une théorie probabiliste de la causalité actuelle. Elle est donc naturellement traitée en revenant à l'idée sur laquelle les théories probabilistes de la causalité sont fondées. Plus précisément, il semble naturel de se demander ce que devient, sous l'interprétation de la conditionalisation que nous venons de proposer, l'idée séminale à partir de laquelle les théories probabilistes de la causalité se sont développées.

6.5.1.1 L'idée séminale sous l'interprétation proposée pour la conditionalisation

L'idée à partir de laquelle les théories probabilistes de la causalité se développent a été présentée dans l'introduction et discutée plus longuement dans la sous-section 2.2.1. Elle consiste à analyser la causalité comme l'augmentation d'une probabilité conditionnelle. Ainsi, il s'agit de considérer que « A cause B » s'analyse fondamentalement au moyen de l'inégalité $Pr(B|A) > Pr(B)$. Nous avons expliqué dans la section 2.2 que cette inégalité est insuffisante à caractériser la causalité, en particulier parce qu'elle ne permet pas de distinguer entre les causes et les effets et parce qu'elle vaut d'effets communs à une même cause. De façon sensiblement différente, nous montrons maintenant que, en-deçà de ces difficultés, l'inégalité $Pr(B|A) > Pr(B)$ ne suffit pas à une analyse même séminale de la causalité singulière. En d'autres termes, l'idée séminale même demande à être complétée quand on passe du cas générique au cas singulier.

L'idée séminale dans le cas singulier. Ainsi que nous venons de l'annoncer, l'ingrédient probabiliste – dont la version la plus rudimentaire est l'idée séminale que nous venons de rappeler – est clairement insuffisant à analyser la causalité quand on abandonne le cas générique pour le cas singulier. En effet, une analyse de la causalité singulière en termes probabilistes ne saurait faire l'économie d'une clause stipulant que l'événement causant et l'événement causé sont effectivement advenus. Ainsi, l'absorption de poison n'est une cause singulière de la mort de Socrate que si a) Socrate a effectivement absorpé du poison et b) Socrate est effectivement mort. A l'inverse, la vérité de l'énoncé « le poison dans la soupe cause (génériquement) la mort » n'implique l'occurrence ni de l'un, ni de l'autre des *relata* causaux. Surtout, elle n'implique pas non plus l'occurrence d'aucun événement singulier qui correspondrait à la cause, ni l'occurrence d'aucun événement singulier qui

correspondrait à l'effet.[70]

Avant de formuler l'idée séminale en tant qu'elle est adaptée au cas singulier, rappelons que, dans les cas tels que $Pr(A) \neq 1$, l'inégalité $Pr(B|A) > Pr(B)$ est équivalente à $Pr(B|A) > Pr(B|\overline{A})$ (voir la proposition 2.2). Cette dernière formulation facilite l'analyse que nous menons dans la présente sous-section ; aussi la retenons-nous. Dans ces conditions, l'idée séminale devient la suivante :

Proposition 6.1 (Idée séminale dans le cas singulier) *A cause (singulièrement) B si :*

1. *(a) l'événement A advient ;*
   (b) l'événement B advient ;
2. $Pr(B|A) > Pr(B|\overline{A})$.

Quel conditionnant fondamental ? La question qui se pose immédiatement à la lecture de la proposition 6.1 est la suivante : quelle est la fonction de probabilités Pr pour laquelle on va se demander ce que devient l'idée séminale sous l'interprétation de la conditionalisation que nous proposons ? En d'autres termes : relativement à quel ensemble de conditions physiques convient-il de définir les probabilités $Pr(B|A)$ et $Pr(B|\overline{A})$?

A ce point, il convient de rappeler que nous avons considéré que les ensembles de conditions physiques relativement auxquels les fonctions de probabilités sont définies dans le contexte propensionniste – les conditionnants fondamentaux – sont composés d'un système physique à un instant donné. Dans ces conditions, la question qui nous occupe dans le présent paragraphe doit être précisée de la façon suivante : relativement à quel système et à quel instant du temps doit-on se demander ce que devient l'idée séminale pour l'analyse de « A cause (singulièrement) B » ?

Pour ce qui est, d'abord, du système, on peut s'en tenir à la caractérisation retenue dans la sous-section 4.1.1 : un système est un ensemble d'objets régis par des mécanismes causaux (au sens générique), relativement isolé et qu'il est pertinent d'analyser pour lui-même. Pour l'analyse de « A cause B », nous considérerons un système en ce sens, qui soit susceptible de produire B. Notons que cette caractérisation est compatible, à la limite, avec la thèse

[70] Cette affirmation peut s'entendre en deux sens. En un sens faible, il s'agit de ceci qu'un énoncé causal générique n'implique l'occurrence d'aucun événement singulier *en particulier* (la mort dans les conditions que nous savons). En ce sens faible, qui suffit à l'argument du présent paragraphe, l'affirmation n'est pas discutable. Mais le sens fort, quant à lui, l'est. Il consiste à affirmer qu'une relation de causalité générique peut exister en l'absence de toute relation de cause à effet singulière qui lui corresponde. Cette thèse est soutenue en particulier par Eells (Eells (1991) pp. 7–8).

selon laquelle les propensions sont relatives à la situation physique globale à un instant donné.

Pour ce qui est, maintenant, de l'instant auquel les probabilités sont définies, la question se présente comme plus difficile. Toutefois, puisque d'une part l'idée séminale consiste à caractériser une cause par ce qu'elle fait à la probabilité de son effet et d'autre part la conditionalisation telle que nous proposons de l'interpréter fait varier les systèmes plutôt que les instants du temps, il convient de faire en sorte que l'instant choisi soit neutre pour l'analyse. C'est le cas si on retient comme pertinent pour l'analyse de « A cause B » l'instant t_A qui caractérise l'événement A. Cela étant posé, nous pouvons en venir à la question se savoir ce que devient, sous l'interprétation de la conditionalisation que nous proposons, l'idée d'analyser la causalité comme une augmentation de probabilité.

Que devient la proposition 6.1 sous notre interprétation de la conditionalisation ? Commençons par la clause 2. Sous l'interprétation que nous proposons pour la conditionalisation, elle revient à ceci : en t_A, $Pr(B)$ est plus élevée dans le système le plus similaire au système qu'on considère parmi ceux relativement auxquels A la probabilité 1, que dans le système le plus similaire au système qu'on considère parmi ceux relativement auxquels la probabilité de $\overline{A}$ vaut 1.

Prenons maintenant en compte la clause 1.(a), selon laquelle A advient en t_A. Cette clause a pour conséquence que le système le plus similaire au système considéré parmi ceux dans lesquels la probabilité de A vaut 1 à l'instant t_A est le système considéré lui-même. Autrement dit, elle a pour conséquence que, en t_A, $Pr(B|A) = Pr(B)$.

Venons en, maintenant, au second membre de l'inégalité. Sous l'interprétation que nous proposons pour la conditionalisation, elle devient $Pr(B)$ à l'instant t_A dans le système le plus similaire au système considéré parmi ceux qui donnent la probabilité 1 à $non - A$. Il s'agit de $Pr(B)$ à l'instant t_A donc du système le plus similaire au système considéré parmi ceux dans lesquels A n'advient pas.

Au total, sous l'interprétation de la conditionalisation que nous proposons, l'idée exprimée par la proposition 6.1 devient la proposition suivante :

Proposition 6.2 (Idée séminale interprétée) *A cause (singulièrement) B si et seulement si :*

1. *(a) A advient ;*
   (b) B advient ;

2'. la probabilité de B en t_A est plus grande dans le système considéré que dans le système qui lui est le plus similaire parmi ceux dans lesquels A n'advient pas.

Il nous semble clair que cette analyse est proche de celle que propose Lewis pour l'analyse de la causalité déterministe (*chancy causation*). C'est sur ce point qu'il convient de faire porter maintenant notre analyse.

6.5.1.2 Comparaison avec l'analyse de Lewis

L'analyse de Lewis. Pour commencer, montrons qu'une analyse de la causalité singulière qui serait constituée des deux clauses 1. et 2'. est effectivement très proche de celle que propose Lewis pour la causalité indéterministe. Pour cela, citons Lewis :

> Mais il y a un second cas [de causalité indéterministe] à considérer : c advient, e a une chance x d'advenir, et effectivement advient ; si c n'était pas advenu, e aurait gardé une chance y d'advenir, mais seulement une chance très mince puisque y aurait été bien inférieur à x. Nous ne pouvons pas vraiment dire que sans la cause, l'effet ne serait pas advenu ; mais nous pouvons dire que sans la cause, l'effet aurait été beaucoup moins probable que ce qu'il était effectivement. Dans ce cas aussi, je crois que nous devrions dire que e dépend causalement de c, et que c est une cause de e.[71]

Le conditionnel contrefactuel : « Si c n'était pas advenu, alors e aurait été beaucoup moins probable » est clairement conçu par Lewis comme requérant une analyse en termes de mondes possibles. En effet, dans l'article consacré à la causalité déterministe, il écrit que les conditionnels contrefactuels qui entrent dans leur analyse[72] doivent être « pris au pied de la lettre : comme des affirmations relatives aux possibilités rivales de la situation actuelle »[73].

La proposition de Lewis est donc la suivante :

Proposition 6.3 (Analyse de Lewis) *A cause (singulièrement) B si et seulement si :*

1. *(a) A advient ;*
   (b) B advient ;

[71] Lewis (1986b) p. 176.

[72] L'analyse est la suivante : A cause B si et seulement si B ne serait pas advenu si B n'était pas advenu.

[73] Lewis (1973) pp. 557–558. Voir aussi p. 560.

2". la probabilité de B est plus grande dans le monde actuel que dans le monde le plus similaire au monde actuel parmi ceux dans lesquel A n'advient pas.

Cette analyse est effectivement très proche de ce que devient l'idée séminale sous l'interprétation de la conditionalisation que nous proposons. Cette proximité est d'autant plus grande que Lewis considère explicitement que les probabilités qui entrent dans l'analyse de la causalité singulière indéterministe sont des probabilités objectives singulières – « single - case chances »[74]. Cette proximité, toutefois, ne saurait masquer certaines différences. Ces différences sont de deux types : relatives aux objets de l'analyse pour certaines, à son contenu pour d'autres. Nous les discutons dans cet ordre.

Comparaison des objets. En premier lieu, il semble exister deux différences entre les objets de la proposition d'analyse 6.2 et ceux de l'analyse de Lewis. Plus précisément, les objets des deux analyses semblent différer par deux aspects. Le premier consiste dans ceci, que nous avons déjà vu, que l'analyse proposée dans Lewis (1986b) vise la causalité indéterministe, là où la proposition 6.1 est une idée pour l'analyse de la causalité – tout court.

Examiner cette première différence requiert de comprendre en quel sens Lewis entend la distinction entre causalité déterministe et causalité indéterministe. C'est Lewis (1973) qui est le plus clair sur ce point :

> Par déterminisme je ne veux pas parler d'une quelconque thèse de nécessitation universelle ou de prédictibilité-en-principe universelle, mais plutôt de ceci : les lois de la nature qui l'emportent sont telles qu'il n'existe pas deux mondes possibles qui sont exactement identiques jusqu'à un moment donné, qui diffèrent ensuite, et dans lesquels ces lois ne sont jamais violées.[75]

Quoi qu'il en soit du détail de la caractérisation de Lewis, ce passage fait apparaître que le déterminisme est conçu comme une propriété du monde, ou plutôt de l'univers auquel appartiennent les mondes possibles – en tout cas pas une propriété des relations de cause à effet. En conséquence, l'analyse de Lewis (1973) et de Lewis (1986b) ne sont pas des analyses de deux espèces d'un même genre, mais l'analyse du même objet (la causalité) sous deux hypothèses différentes (déterministe et indéterministe).

Or, l'introduction de probabilités dans l'analyse de la causalité découle précisément de l'affirmation selon laquelle la causalité n'est pas une relation de nécessitation. En d'autres termes, les théories probabilistes de la causalité

[74]Lewis (1986b) p. 177.
[75]Lewis (1973) p. 559.

n'ont de sens que sous la seconde de deux hypothèses envisagées par Lewis. Dans ces conditions, l'analyse proposée dans Lewis (1986b) et l'analyse de la causalité qui découle de la proposition 6.1 sous l'interprétation avancée pour la conditionalisation ont bien le même objet : la causalité indéterministe au sens que Lewis donne à ce terme.

Le second point par où l'objet de la proposition d'analyse 6.2 diffère de celui de l'analyse menée dans Lewis (1986b) est le suivant : de même que les probabilités quand on en donne une interprétation propensionniste, les relations de cause à effet telles qu'elles sont analysées par la proposition 6.2 sont relatives à un système physique. A l'inverse, l'analyse de Lewis vise la causalité en sens absolu. Nous avons déjà rencontré une situation de ce type dans le chapitre 2 : alors que les réseaux bayésiens causaux véhiculaient une caractérisation de la causalité directe au sein d'un ensemble de variables, les théories probabilistes de la causalité visent à analyser une concept absolu de causalité. Il est alors apparu que la causalité absolue au sens du chapitre 2 pouvait être considérée comme la causalité relative à un ensemble de variables suffisant à décrire la réalité empirique. De façon analogue, la causalité absolue au sens du présent chapitre peut être considérée comme un cas particulier limite de causalité relative. Plus spécifiquement, la causalité non relative à un système physique peut être considérée comme la causalité relative à ce système physique particulier qu'est la situation physique globale. L'objet de l'analyse 6.2 et l'objet de l'analyse menée de Lewis (1973) coïncident sous l'hypothèse selon laquelle les conditionnants fondamentaux des fonctions de probabilités interprétées de façon propensionniste sont la situation physique globale à un moment donné. En vue d'autoriser la comparaison des deux analyses, nous émettons cette hypothèse et nous y tenons jusqu'à la fin de la présente sous-section.

Comparaison du contenu des analyses. Pour ce qui est des analyses aussi, nous identifions deux différences entre la proposition d'analyse 6.2 et l'analyse de Lewis. La première concerne l'instant du temps auquel les probabilités qui interviennent dans l'analyse de la causalité singulière sont définies. Nous avons considéré qu'elles devaient être des probabilités à l'instant t_A caractéristique de l'événement-cause A. De son côté, Lewis considère les probabilités à l'instant immédiatement postérieur à t_A.[76] Notons que Lewis ne justifie pas cette décision ; il nous semble qu'elle procède du projet de capturer exactement ce que l'occurrence de A fait à B – ou, plus précisément ici, à la probabilité de B. Or, nous ne voyons pas ce qui impose d'attendre l'instant postérieur à t_A pour mener ce projet à bien ni, du coup, pourquoi

[76]Lewis (1986b) pp. 176–177.

ce n'est pas en considérant la situation en t_A qu'on saisit avec le plus de justesse ce que l'analyse prétend saisir. Nous soutenons donc à la fois que la divergence dont il est question ici est minime et que, à choisir, il nous semble préférable de considérer les probabilités en t_A.

Pour ce qui est, maintenant, de la seconde différence entre la proposition d'analyse 6.2 et l'analyse de Lewis, elle est la suivante. D'un côté, la proposition 6.2, et avant elle l'idée sur laquelle les théories probabilistes sont fondées, consiste à analyser la causalité simplement comme une augmentation de probabilité. De l'autre côté, pour Lewis, A ne cause B que si B aurait été « *beaucoup moins* probable » (*much less probable*) en l'absence de A qu'il ne l'est en sa présence. Lewis propose donc une analyse de la causalité qui est plus exigeante, mais moins précise que l'analyse exprimée par la proposition 6.2. En effet, s'il indique que « beaucoup moins » telle qu'il fait occurrence dans l'analyse de la causalité signifie « moins par un grand rapport – non par une grande différence »[77], Lewis ne fixe pas une valeur à partir de laquelle un rapport de probabilité est suffisamment grand pour qu'il y ait causalité. Surtout, nous ne voyons pas quelle pourrait être une telle valeur et selon quels critères elle pourrait être fixée. Si elle est plus exigeante, l'analyse de Lewis est donc effectivement plus vague. Maintenant, de manière plus générale, la différence dont nous faisons état dans le présent paragraphe apparaît comme la seule différence substantielle entre l'analyse de la proposition 6.2 et l'analyse développée dans Lewis (1973).

En définitive, nous avons montré que l'analyse développée dans Lewis (1973) et l'analyse de la proposition 6.2 n'ont pas des objets disjoints. Plus précisément, il est apparu que l'objet de la première analyse peut être considéré comme un cas particulier de l'objet plus générique de la seconde analyse. Dans ces conditions, il nous a été possible de comparer les analyses elles-mêmes – c'est-à-dire leur contenu, et non seulement leurs objets. Une seule différence substantielle apparaît alors, qui est précisément celle qui existe entre « beaucoup plus probable » et « plus probable ». L'interprétation de la conditionalisation que nous avons proposée dans la section 6.3 est donc telle qu'elle rend similaires l'idée sur laquelle les théories probabilistes de la causalité sont fondées et l'analyse de la causalité par Lewis. Autrement dit, moyennant l'interprétation proposée pour la conditionalisation et le fait qu'on s'intéresse à la causalité singulière, l'idée d'analyser la causalité comme une augmentation de probabilité conditionnelle coïncide avec l'analyse que Lewis donne de la causalité indéterministe dans Lewis (1986b). En ce sens, notre proposition d'interprétation propensionniste de la conditionalisation

[77] Lewis (1986b) p. 177.

contribue à l'analyse probabiliste du concept de causalité singulière – et, plus précisément ici, actuelle.

6.5.2 Deux notions de causalité

Ce qui a été exposé dans la dernière sous-section n'a de sens que sous l'interprétation de la conditionalisation que nous avons proposée. A l'inverse, la sous-section qui commence vise à apporter des éléments d'analyse du rapport entre propensionnisme et causalité qui ne dépendent pas de l'interprétation proposée pour la conditionalisation. Plus précisément, ces éléments d'analyse s'organisent autour d'une idée principale : le propensionnisme en général, et la discussion du paradoxe de Humphreys en particulier, nous invitent à distinguer entre deux notions de causalité. Pour le comprendre, revenons à la sous-section 6.2.2, et à l'analyse que nous y proposions du désaccord entre Humphreys et McCurdy en tant qu'il porte sur quelque chose de plus profond que l'évaluation d'une probabilité conditionnelle inverse.

Dans la sous-section 6.2.2, il est apparu que le paradoxe de Humphreys repose largement sur l'idée selon laquelle le propensionnisme engage à penser en termes dispositionnels le rapport entre le conditionnant et le conditionné d'une probabilité conditionnelle. Contre cette thèse, nous avons tenu ferme l'idée selon laquelle le propensionnisme engage à interpréter comme tels les seuls rapports entre les ensembles de conditions physiques qui sont les conditionnants fondamentaux des fonctions de probabilités et les événements que ces ensembles de conditions physiques tendent à réaliser. Cette position s'adosse à la caractérisation du propensionnisme que nous avons proposée dans le chapitre 5 et permet d'envisager de résoudre le paradoxe. Dans le cadre d'une conception réaliste des propensions, du type de celle qui a été présentée dans le chapitre 5, il s'ensuit que le propensionnisme engage à penser causalement non pas le rapport entre le conditionné et le conditionnant d'une probabilité conditionnelle, mais le rapport entre un ensemble de conditions physiques et un événement.

Toutefois cela n'implique pas qu'il est impossible de penser – comme il est plus commun et comme le suppose l'approche par les théories probabilistes – qu'un *événement* en cause un autre. Il nous semble qu'un des mérites de la proposition d'interprétation de la conditionalisation que nous avons formulée est de le faire apparaître. En effet, nous avons montré dans la dernière sous-section que, sous cette interprétation de la conditionalisation, la relation d'augmentation de probabilité s'entend en un sens très proche de celui où la causalité indéterministe entre événements singuliers est définie dans Lewis (1986b). Il apparaît alors à la fois que le propensionnisme n'interdit pas de penser la causalité entre événements singuliers et que la causalité entre

événements singuliers demande à être analysée dans le contexte propensionniste, qu'elle n'y est pas un concept primitif.

Maintenant, la distinction qui se fait jour nous semble correspondre à celle qui est défendue dans Hall (2001). Dans ce texte, Hall soutient qu'il est nécessaire de distinguer entre deux concepts de causalité, qui ont des propriétés incompatibles :

- la *dépendance* est la relation que capturent les analyses contrefactuelles de la causalité. Il nous semble qu'elle est aussi l'aspect du concept de causalité dont les théories probabilistes visent à rendre compte. Hall la caractérise par ceci qu'elle autorise la causalité d'omissions et par omission ;
- la *production* ne les autorise pas. Mais elle a les trois propriétés suivantes que n'a pas la dépendance :
  1. elle est transitive ;
  2. elle est locale, au sens où « les causes sont connectées à leurs effets *via* des suites spatio-temporellement continues d'intermédiaires causaux »[78] ;
  3. elle est intrinsèque : « la structure causale d'un processus est déterminée par son caractère intrinsèque, non causal (et par les lois) »[79].

Nous ne nous lançons pas dans une analyse de la distinction de Hall, ni surtout dans une discussion des voies difficiles qu'il emprunte pour l'asseoir. Simplement, nous aimerions suggérer que la notion de causalité à laquelle le propensionnisme réaliste engage est une notion de production. D'un côté, en effet, la production se présente comme 2. un engendrement au sens physique du terme, qui 3. dépend de ce qu'est la situation et des lois de la nature. Or, dans un contexte propensionniste, la relation entre les conditionnants fondamentaux et les événements singuliers est précisément d'engendrement et déterminée (pour autant qu'elle l'est) en termes physiques non causaux.

De l'autre côté, le fait que la relation de production définie par Hall est une relation entre deux événements singuliers – et non entre un ensemble de conditions physiques et un événement – n'implique pas que la relation entre conditionnants fondamentaux et arguments des fonctions de probabilités propensionnistes n'est pas de production. En effet, Hall fait dériver la relation de production entre événements d'une relation plus fondamentale, entre des ensembles minimaux d'événements et les événements qu'il suffisent à produire[80]. Plus précisément, les antécédents de la relation de production étant

[78] Hall (2001) p. 225.
[79] Hall (2001) p.225.
[80] Hall (2001) p. 259–260.

les événements qui appartiennent à de tels ensembles. Etant donné ce qui a été dit plus haut dans le présent paragraphe, la relation entre ensembles minimaux d'événements et événements qu'ils suffisent à produire apparaît être du même type que la relation d'engendrement d'un événement singulier par un ensemble de conditions physiques.

Entrer dans le détail de la comparaison entre les deux relations que nous venons de mentionner dépasse le cadre du présent travail. Il nous semble toutefois que nous avons suffisamment justifié l'idée selon laquelle la notion de causalité qui est éventuellement primitive dans le propensionnisme est du même type que la production telle que Hall la définit. Si c'est bien le cas, il apparaît que le rapport entre la causalité et les probabilités est différent selon qu'on envisage la causalité comme production ou comme dépendance. Pour ce qui est de la causalité comme production, le rapport le plus étroit – que nous avons envisagé prioritairement dans le chapitre 5 – est de recours à la causalité comme un *concept primitif* dans l'interprétation des probabilités objectives singulières. Pour ce qui est de la causalité comme dépendance, elle est *ce qu'on cherche à analyser* dans le cadre des théories probabilistes de la causalité. Cela va aussi bien dans le cas générique qui nous a occupé dans la première partie de notre travail, que dans le cas singulier. La question abordée dans la dernière sous-section (6.5.1) se présente alors comme celle l'analyse de la dépendance entre événements singuliers au moyen de probabilités conçues comme faisant référence à des relations de production.

Conclusion

Dans le travail qui s'achève, nous avons traité de questions que soulèvent les théories probabilistes de la causalité dans l'état actuel de leur développement. Ainsi que nous l'avons expliqué très tôt, ces questions sont différentes selon qu'on s'intéresse à la causalité générique – entre des propriétés – ou à la causalité singulière – entre des événements. Pour terminer, nous rappelons les résultats obtenus concernant l'une et l'autre, en même temps que nous en précisons la portée.

En ce qui concerne la causalité générique, ce sont des questions d'*épistémologie* que nous avons abordées. En d'autres termes, la première partie de notre travail a porté sur l'identification des relations de cause à effet entre propriétés. Cette question a été posée comme une question relative aux réseaux bayésiens. En effet, ceux-ci se sont présentés d'emblée comme les candidats les plus naturels au titre d'outils d'inférence causale corrélatifs des théories probabilistes de la causalité. Il convient de souligner que c'est seulement en tant que tels que les réseaux bayésiens ont été envisagés. En particulier, les questions soulevées par l'utilisation des réseaux bayésiens pour modéliser la connaissance incertaine dépassent clairement le cadre de l'enquête qui s'achève.

L'idée selon laquelle les réseaux bayésiens sont les outils d'inférence causale corrélatifs des théories probabilistes de la causalité a été avérée par le chapitre 2. D'un côté, nous avons montré que le critère de causalité véhiculé par les réseaux bayésiens causaux s'inscrit de plain-pied dans le champ des théories probabilistes de la causalité. De l'autre côté, nous avons montré que ce qui sépare ce critère de nos meilleures théories probabilistes de la causalité suffit à le rendre utilisable.

Concernant précisément l'utilisation des réseaux bayésiens pour inférer des causes, on peut considérer comme fondamentale la thèse soutenue à la fin du chapitre 1, selon laquelle les hypothèses relatives à la causalité et à son rapport avec les probabilités sont substantielles dans le contexte d'*inférence* causale. Par là, nous entendons exactement que, dans le contexte d'inférence causale, il n'est pas possible, quand ces hypothèses sont violées, de les rétablir

en amendant localement le graphe interprété causalement.

De cela il découle en premier lieu qu'il est légitime de se demander quelles sont les limites du domaine au sein duquel ces hypothèses sont satisfaites. Plus précisément, la tâche telle qu'elle s'est spécifiée consiste à caractériser des classes de systèmes qui soit satisfont prouvablement ces hypothèses, soit échouent prouvablement à les satisfaire. Dans le chapitre 4, nous avons discuté un argument de Steel visant à montrer que le caractère déterministe n'est pas discriminant pour ce qui est de la satisfaction de l'hypothèse centrale qu'est la condition de Markov causale. Contre lui, nous avons défini un sens précis auquel les systèmes déterministes sont plus susceptibles que les systèmes indéterministes de satisfaire cette condition.

En second lieu, le fait que les hypothèses véhiculées par les réseaux bayésiens causaux sont substantielles dans le contexte d'inférence causale a des conséquences importantes du point de vue de la méthodologie de l'inférence causale. Plus précisément, les conclusions du chapitre 3 dépendent très largement de ce que ces hypothèses sont substantielles en même temps que susceptibles d'être violées. En effet, à partir d'une analyse de ce qui découle de la possibilité que ces hypothèses substantielles soient violées, nous avons soutenu que les résultats de l'inférence causale fondée sur les réseaux bayésiens sont *toujours* suspects de ne pas être corrects. Il en découle que les réseaux bayésiens ne peuvent pas nous tenir lieu de seuls outils d'inférence causale. Utiliser les réseaux bayésiens pour inférer des causes devient alors leur ménager une place au sein de nos procédures traditionnelles d'inférence causale. Nous avons indiqué selon quelles voies cela nous semble pouvoir être fait.

Du côté de la causalité singulière, ce ne sont pas des questions d'épistémologie qui nous ont intéressés. En effet, nous avons montré que, en amont des questions d'épistémologie, une théorie probabiliste du concept de cause actuelle reste à formuler. C'est dans la voie ainsi ouverte que nous nous sommes engagés dans la seconde partie. Les résultats principaux sont venus tard, pour des raisons que nous avons déjà dites mais que nous rappelons une dernière fois ici. D'une part, une caractéristique essentielle des théories probabilistes de la causalité est qu'elles recourent aux probabilités conditionnelles. D'autre part, le projet de théorie probabiliste de la causalité actuelle n'a de sens que si les probabilités reçoivent une interprétation propensionniste. Dans ces conditions, résoudre le problème que le propensionnisme rencontre au moment d'interpréter les probabilités conditionnelles devient un préalable à l'analyse.

Dans le chapitre 6, nous avons formulé une proposition d'interprétation propensionniste de la conditionalisation. En tant que telle, cette proposition relève de la philosophie des probabilités. En tant qu'elle prend place dans

notre travail, elle fonde deux analyses conclusives. Ainsi il est apparu que l'interprétation proposée pour la conditionalisation est telle que la notion d'augmentation de probabilité conditionnelle devient similaire à l'analyse de Lewis pour la causalité actuelle[81] dans le cas indéterministe. Pour le dire plus clairement : *modulo* l'interprétation proposée pour la conditionalisation, l'idée qu'on trouve au fondement des théories probabilistes de la causalité coïncide pratiquement avec l'analyse lewisienne pour la causalité actuelle.

Enfin, l'extension du propensionnisme aux probabilités conditionnelles a rendu urgente la distinction entre deux notions de causalité. Cette distinction n'est pas nouvelle, mais nous semble particulièrement éclairante quand on l'introduit dans le champ des discussions sur le rapport entre causalité et probabilités, et particulièrement quand on la rapporte au propensionnisme. La première notion de causalité est alors celle qui est à l'oeuvre quand on dit que le propensionnisme popperien tel que nous l'avons présenté dans le chapitre 5 est une interprétation causale des probabilités. La causalité est alors production, ici production d'un événement par un ensemble de conditions physiques. La seconde notion de causalité est la notion de dépendance qu'on caractérise comme une augmentation de probabilité conditionnelle. Ainsi, de façon plus générale, la dépendance causale est ce que les théories probabilistes de la causalité visent à caractériser.

[81] Lewis ne parle pas de « causalité actuelle », mais plus simplement de « causalité singulière ». Toutefois le sens qu'il donne à cette expression est celui que nous donnons à « causalité actuelle ».

Annexes

Note on the appendices

The following appendices are written in English and should facilitate the reading of the French part of the dissertation.

Appendix A is made up of three articles that are related to my dissertation. The content of these papers has been integrated into a different chapter of the dissertation. Except from minor corrections, the papers are included here in their original form. This means, however, that the corresponding chapters contain more material than the papers. I hereby indicate how to make the correspondence.

Paper A.1, "Causal inference: How can Bayes nets contribute?" (Drouet (2007)), was written on the occasion of the "Causality and probability in the sciences" conference that took place in Canterbury in June 2006. The question tackled in this paper is the same as the one tackled in chapter 3 of the dissertation. The main argument is the same in both cases. However, the analyses in chapter 3 are noticeably more detailed and qualified as the one in the paper.

Paper A.2, "Is determinism more favorable than indeterminism for the causal Markov condition?", was written on the occasion of a "Philosophy, Probability, Physics" workshop that took place in Paris in April 2006. It is unpublished to date. Its content largely coincides with the one of chapter 4 of the dissertation. The only noticeable difference consists in the chapter containing a more complete discussion of the notion of determinism that appears in discussions concerning Bayesian networks.

Paper A.3, "Can there be a propensity interpretation of conditional probabilities?", was written on the occasion of the "Reasoning about probabilities and probabilistic reasoning" that took place in Amsterdam in May 2007. It is unpublished, but is under evaluation. Its content is integrated with chapter 6 of the dissertation. However the chapter is much richer than the paper. On the one hand the analysis of Humphreys paradox and on the other hand the consequences of the way I propose to solve (specifically the one that are relative to the relationship between causality and probability) are explored at greater length than in the paper.

Appendix B is an extended abstract of the dissertation. More precisely, the abstract begins with a translation of the totality of the French introduction and goes on with an abstract of each of the chapters, followed by an abstract of the conclusion. Although less detailed, the outline of the abstract exactly matches that of the French part of the dissertation.

Annexe A

Articles en anglais

A.1 Causal inference: How can Bayes nets contribute?

Abstract

Inference of causal knowledge from statistical data is a very old problem. Yet, for twenty years now, Bayes nets algorithms seem to have considerably renewed it. Bayes nets algorithms have been extensively and heatedly debated. A fundamental criticism deals with the relationship they assume between causality and probability. What emerges from these debates is that some systems satisfy this relationship while others do not. On this basis, the paper aims at assessing the contribution of Bayes algorithms to causal inference independently from the problem raised by Bayes nets assumptions. The paper begins with a comparison between Bayes nets causal inference and traditional causal inference for social sciences systems satisfying Bayes nets assumptions. The conclusion is very positive: for these systems, Bayes nets causal inference well and truly has many advantages over the traditional methodology it competes with. In the second part of the paper, I go back over the restriction to systems satisfying Bayes nets assumptions and show those systems cannot be identified before causes are known. This leads me to propose a mixed methodology for causal inference which enables to make use of Bayes nets algorithms in spite of the previously highlighted difficulty.

A.1.1 Introduction

Whether it is possible to infer causal knowledge from observational data and how to do it are very old questions for both philosophers and scientists. It

has been thought since the beginning of the 1990s that Bayesian networks might provide new answers for those old questions. More precisely, various algorithms based on Bayesian networks have been introduced that are said to output causal knowledge from observational probabilistic information.[1] These algorithms have given rise to a heated debate, fed by papers ranging from the most enthusiastic support to the harshest criticism.[2] This debate has very quickly focused (and still focuses) on the assumptions conveyed by Bayes nets algorithms about the relationship between causality and probability. The discussion concerning those assumptions is obviously not over; yet it has already made it clear that some systems satisfy Bayes nets assumptions whereas other ones do not.

The present paper aims at assessing how Bayes nets algorithms can actually contribute to causal inference. This question is important from a practical point of view: one would like to know whether Bayes nets causal inference algorithms can be used, when and how they can be used, and what can be expected from this use. Yet this question is hardly treated independently from the heated discussion about the validity of Bayes nets assumptions. On the one hand, that Bayes nets assumptions hold in most interesting cases is a key claim of the proponents of Bayes nets causal inference (see [Korb97] p. 547, [Pearl00]pp. 61-63, [Schei97]pp. 190-194, [SGS]pp. 32-42). On the other hand, that Bayes nets assumptions may be violated is a key argument of those who are skeptical about Bayes nets causal inference (see [Cart99], [Cart01], [Free99]pp. 31-34). I would like to slightly shift the emphasis of the debate, and to provide an assessment of Bayes nets algorithms that does not depend on how the possible violations of Bayes nets assumptions are interpreted – but just takes them for granted. The paper relies on the claim that taking those possible violations for granted gives means to broach the question of the contribution of Bayes nets algorithms to causal inference. Indeed, knowing that there exist systems satisfying Bayes nets assumptions allows to focus on those specific cases, and knowing that Bayes nets assumptions do not hold universally enables to identify as essential the question of what the consequences of this non-universality are. The outline of the paper accords with those two remarks. First I restrict the analysis to systems satisfying Bayes nets assumptions; for these systems, the assessment is very favourable to Bayes nets algorithms. Second, I go back to the general *via* an analysis of the significance of this restriction. The analysis shows that the existence of systems not satisfying Bayes nets assumptions precludes rough Bayes nets causal inference being performed in any real case, and leads me to propose

[1] A good and still recent review on these algorithms is [SGS] chap. 5 and 6.

[2] [CinC] gives a significant sample.

the integration of Bayes nets algorithms to a wider mixed methodology for causal inference. The whole analysis is restricted to the social sciences – the domain which Bayes nets algorithms most directly address[3] –, and to cases without latent variables. These restrictions enable to keep very tractable from a technical point of view while already broaching essential philosophical issues.

A.1.2 Systems satisfying Bayes nets assumptions

As already stated, this section focuses on systems satisfying "Bayes nets assumptions". Those assumptions do not deal directly with systems (be they physical, economical, biological...), but rather with the sets of variables representing them. For such a set **V**, Bayes nets assumptions are the following ones:

- **Acyclicity**: the directed graph whose arrows represent direct causal relations among **V** variables is acyclic;
- **Causal Markov Condition**: any variable in **V** is probabilistically independent from all variables in **V** but its effects when one conditionalizes on its direct causes;
- **Faithfulness**: there are no probabilistic independencies among the variables in **V** but the ones that are entailed by the Causal Markov Condition.

I consider those Bayes nets algorithms that are most discussed by philosophers, that is algorithms that infer causal knowledge from probabilistic independencies[4]. For cases without latent variables, the best-known amongst those algorithms are the IC algorithm introduced by Pearl and Verma (for a presentation, see [Pearl00]pp. 50-51) and the PC algorithm introduced by Spirtes, Glymour and Scheines (see [SGS]pp. 84-88). Once a set **V** of variables has been selected in order to represent the system under consideration[5], those algorithms perform causal inference in three steps:

1. measure the values of the variables in **V** for the members of a representative and large enough sample of the population you are interested in;

[3]The reasons for this are that Bayes nets algorithms deal with causal inference in non-experimental contexts and that the impossibility of experiments is characteristic of the social sciences.

[4]In particular, I will not discuss works dealing with learning Bayesian networks using Bayesian techniques.

[5]As already stated, **V** is assumed to be causally sufficient.

2. compute the standardized correlations, absolute and partial, among **V** variables;

3. test each of the computed correlations against the null hypothesis that it is zero and identify the set **I** of correlations for which the null hypothesis cannot be rejected;

4. construct a partially directed graph over **V** that represents all the causal knowledge that can be inferred from **I** owing to Bayes nets assumptions concerning the relationship between null partial correlations and direct causal relations. This graph is the output of the procedure. It represents (all and only) the binary causal relations shared by the causal structures compatible with the data plus Bayes nets assumptions.

The methodology I take up in order to assess Bayes nets causal inference when Bayes nets assumptions are satisfied is comparative. Therefore, the question is now: which methods for causal inference in the social sciences should Bayes nets algorithms be compared with? The most obvious answer is that they should be compared with automated structure learning procedures they directly compete with, in particular those operationalized by the LISREL and EQS packages. There are two reasons why the present comparison will not be with procedures of that kind. First, Spirtes, Glymour and Scheines have given several convincing reasons why Bayes nets algorithms give more satisfactory results than those procedures.[6] Second, automated search procedures are scarcely used by social scientists, who still massively resort to more traditional techniques. Therefore, if one is to determine whether and how Bayes nets algorithms can contribute to actual causal inference, these algorithms should be compared with those more traditional techniques – which has not been done yet. It is true that this comparison raises methodological difficulties. I will discuss them as they emerge, and try to show that they can be overcome.

The "more traditional techniques" I will compare Bayes nets algorithms with stem from path analysis. The way social scientists actually use those techniques can and must be criticized. Yet this has already been extensively done by both opponents (see for instance [Free91]) and proponents (see [Kli98] chap. 12), and they widely agree on the main misuses of path analytic causal inference techniques, some of the reasons for them (on this point, see in particular [Bla91]) and the ways they can be avoided. As a consequence, the present paper focuses on path analytic causal inference as it

[6] [SGS]pp. 74-80.

should be performed in order to give its bets results. Accordingly, I consider a procedure which is not standard, but rather is a construct which I think is the most rigorous and complete methodological proposition that can be extracted from usual social science practice. Once a set **V** of variables representing the system under consideration has been identified, the procedure runs as follows:

1'. specify a model **M** that you think might adequately represent the causal relationship among **V** variables. As already stated, it is assumed that **M** is a directed acyclic graph over **V**, whose arrows (or "paths") represent hypothesized direct causal relations among the variables in **V**. As far as possible, **M** should be overidentified by the data that will be collected;

2'. measure the values of the variables in **V** for the members of a representative and large enough sample of the population you are interested in, and compute the standardized correlations among **V** variables;

3'. use the data collected at step 2. in order to estimate the structural coefficients in **M**. Each of these coefficients evaluates the strength of exactly one of the arrows in **M**;

4'. test **M'**, that is:

- determine whether the causal relationship and the signs and absolute values of structural coefficients look plausible;
- if the model is overidentified, compute the correlation residuals, that is the differences between model-implied and observed correlations and check that none has absolute value greater than .1;[7]
- if the model is overidentified, identify the independent overidentifying restrictions, state them as null hypotheses, test these null hypotheses and check that they cannot be rejected;

5'. if **M'** does not pass these tests, reject it, come back to 1. and specify another model. If it does pass them, perform the following steps;

6'. if **M'** is overidentified, assess its fit to the data. Many model fit indexes are available and several of them must be performed;

7'. reiterate steps 1'. to 6'. for another model;

[7].1 is a conventional but generally accepted maximum for the absolute values of correlation residuals in an acceptable model.

8'. compare the fit scores of the models that have been considered and were not rejected at step 5'. Identify the model **M*** that seems to be the overall better data fitting;

9'. identify models equivalent to **M*** that look plausible representations of the causal structure over **V**, and provide a theoretical justification of your preference for one of them. This model is the output of the procedure.

The main difference between the two procedures that I have just detailed is a methodological one. More precisely, Bayes nets causal inference is deductive whereas path analytic one is not. The case of Bayes nets is the simplest: Bayes nets algorithms output a "pattern" representing all the binary causal relations that deductively stem from the analysis of the data (2. and 3.) under Bayes nets assumptions concerning the relationship between causality and probability. The case of path analytic causal inference as I have presented it is a little more intricate. From a methodological point of view, it is clearly composed of two parts – each of which is autonomous enough to provide the methodology for many social sciences papers. The first part, from 1'. to 5'., can be described as hypothetico-deductive. A hypothesis is formulated as a model is specified (1'.), then consequences are deduced from it in the form of the estimation of structural coefficients (3'.), and finally those consequences are tested against the data (4'. and 5'.). If the hypothesis does not pass the test, it is rejected as refuted; if it does, it is retained as corroborated (end of 5'.). It could be argued that path analytic causal inference has only the appearance of being hypothetico-deductive, and that this appearance badly hides crucial differences between the methodology of path analytic causal inference and standard hypothetico-deduction. In particular, the derivation of consequences from the initial hypothesis does not require auxiliary theoretical hypotheses but data. Moreover, the conclusion is not deductively drawn but only estimated from the premises. Similarly, the rejection of a hypothesis does not rely on a formal contradiction with the data, but on a probabilistic incompatibility. I would answer that the last two differences derive from the statistical nature of the data, but do not matter much at the level of methodological abstraction from which I am currently looking at things. The second part of path analytic causal inference is constituted by 8'. and 9'.. It consists in identifying first the models that best fit the data among the ones that have not been rejected, and second the most plausible among the non rejected models that best explain the data. As fit to data corresponds to the capacity of explaining it, 8'. can be considered as a form of inference to the best explanation. More generically, both 8'. and 9'. aim at selecting a hypothesis among competing ones. That is a matter

of induction, though not of traditional enumerative induction. This kind of induction is described as "hypothetical" by Harman[8].

It is my contention that this methodological divergence results in (at least) three differences which all contribute to the superiority of Bayes nets algorithms over the path analytic procedure that has been described[9]:
First, the procedure resorting to Bayesian networks is data-based, whereas the path analytic one is model-based. Data-baseness is an immediate consequence of deductiveness, together with the fact that analyzed data constitute the premises of the deduction. Similarly, model-baseness is a characteristic common to hypothetico-deduction and hypothetical induction: one has to formulate a hypothesis before its consequences can be tested against the data, as one has to identify competing models before the best of them can be looked for. Model-baseness leads to the dependence of the final output on the causal order the scientist has been able to envisage, and that is clearly a drawback to path analytic causal inference over Bayes nets one. It could be argued that the difference between data and model-baseness is so essential from a methodological point of view that it makes the comparison unfair from the beginning. From that point of view, if Bayes nets algorithms are to be compared with a path analytic procedure, it must be with a procedure that runs through all possible models. Yet such a procedure would clearly be computationally more demanding than Bayes nets algorithms, which are already criticized for requiring the computation of all correlations among the **V** variables. More important, that would deprive path analytic methods from one of their essential features, and obscure the difference between them and the automated search procedures that I decided not to consider. Therefore, I will persist in considering a model-based path analytic procedure, and this model-baseness will count as a remarkable difference between path analytic and Bayes nets causal inferences. Obviously, this difference is in favour of Bayes nets causal inference. Relying on the specification of particular models (step 1'.) is a feature of path analytic causal inference that has been criticized.[10]
Second, Bayes nets procedure outputs a pattern, whereas the path analytic one outputs a single structure. As already stated, the pattern output by Bayes nets algorithms represent the binary causal relations that deductively

[8][Har92].

[9]Note that claiming that Bayes nets causal inference is "superior" to path analytic one does not amount to renouncing to the approach that was assumed at the outset of the paper. Indeed, this approach does not consist of eschewing any kind of evaluation of Bayes nets algorithms, but of evaluating them *independently from the debate on Bayes nets assumptions*. This is clearly what is going on at the present point of the paper.

[10][Free87]pp. 120-121 and [Free91]pp. 303-304 and p. 309.

stem from the analyzed data under Bayes nets assumptions concerning the relationship between causality and probability. Therefore the output of a pattern well and truly derives from deductiveness. Now this pattern stands for the whole class of acyclic causal structures sharing the binary causal relations it depicts; those structures that cannot be distinguished by data under the assumed relationship between causality and probability. By contrast, the output of the path analytic procedure is a causal structure, for the simple reasons that the hypotheses formulated at 2'. are causal structures too and that hypothetical induction aims at isolating **the** hypothesis to be preferred. It is true that step 9'. requires the researcher to identify models that are equivalent[11] to the better-fitting considered model, to decide whether some of them are plausible and to provide a justification for preferring one of them to the others. Yet the realization of latter two tasks depends on both the state of scientific knowledge at the time causal inference is performed and some of the scientist's convictions. Therefore, the nature of its output must be considered an advantage of Bayes nets causal inference over path analytic one.

Third and lastly, and as another immediate consequence of deductiveness, the binary causal relations output by Bayes nets algorithms must be accepted as true as soon as those zero correlations are accepted as true. On the contrary, even one who accepts estimations of structural parameters in 3'. and identifications of the vanishing ones in 4'., has no reason to believe that the structure output by the path analytic procedure adequately represents reality. On the one hand, some structures that have not been considered may better fit data than the output one; on the other hand, there is no guarantee that best-fitting data structures are causally significant. More generally, non demonstrativeness is a recognized feature of both hypothetico-deduction and hypothetical induction.

In this section, I have detailed how causal inference is performed by Bayes nets algorithms on the one hand and by usual path analytic methods on the other hand. I have pin-pointed three features that make Bayes nets causal inference superior to path analytic one. Finally I have shown that those features are correlates of deductiveness, as opposed to the combination of hypothetico-deduction and hypothetical induction in the path analytic case. But the whole analysis concerns only those systems that satisfy Bayes nets assumptions. In the next section, I analyze the impact of this restriction on the conclusions I have drawn and come back to the question of the contribu-

[11] Equivalence of models in the path analytic context concerns the ability to explain observed correlations. It must not be mixed up with indistinguishability in the Bayes nets context, which corresponds to the depiction of the same probabilistic independencies.

tion of Bayes nets algorithms to causal inference in the general case.

A.1.3 General case

At first sight, the restriction of the preceding analysis to cases satisfying Bayes nets assumptions does not raise particular difficulties. First, it looks fair to focus on systems liable to Bayes nets causal inference when examining the contribution of Bayes nets to causal inference. Second, this focus does not seem to have consequences other than the following ones: the conclusions are valid only for the systems that have been taken into account, and only these systems can benefit from the nice features of Bayes nets causal inference that have been highlighted.

The matter is that this first-sight analysis is misleading. What is at stake is not only that Bayes nets assumptions fail to universally hold, but also that we do not know when they do. Let me explain this idea focusing on the Causal Markov Condition ("CMC" from now onwards). There are two justifications for this focus. On the one hand, it is probably more central to inference of causality from probability than the other two Bayes nets assumptions. Correspondingly, it has been much more discussed and is better-known. On the other hand, this focus does not make any difference as far as the conclusions of the paper are concerned. Now it is true that we have a sufficient condition for the satisfaction of the CMC. More precisely, it is now well-known that the CMC holds for deterministic systems represented by functional models with independent exogenous variables[12]. In other words, given a system and a set of variables $\mathbf{V}$ causally sufficient for this system, $\mathbf{V}$ satisfies the CMC if

a. the value of any $\mathbf{V}$ variable is functionally determined by the values of its direct causes in $\mathbf{V}$;

b. any two disjoint non-empty subsets of the set of $\mathbf{V}$ variables that do not have any causes in $\mathbf{V}$ are probabilistically independent.

Although a sufficient condition for the satisfaction of the CMC, this cannot work as a criterion in the context of causal inference. Indeed, the very definitions of functional models and of exogenous variables imply that these notions are meaningful only for one who knows of the causal structure of the system

[12][Pearl00]p. 30 and [SGS]p. 32. The result also holds for pseudo-deterministic ([SGS]pp. 27-28.) and indeterministic ([Steel05]) systems. These extensions involve latent variables, hence they will not be taken into account in the present paper. Nevertheless, their existence constitutes an argument in favour of the possibility to extent the results of the paper beyond the scope of the restrictions which were imposed at the output.

being considered. But that – knowledge of the causal structure – is exactly what we are looking for when performing causal inference. Furthermore, I do not know of any non causal criterion for the satisfaction of the CMC[13]. As a consequence, it is impossible to know whether Bayes nets algorithms can be rightfully used before causal inference is actually performed.

There exist arguments to the effect that Bayes nets algorithms should nevertheless be used each time no violation of Bayes nets assumptions has been detected beforehand.[14] Yet I think that no argument of this kind may be accepted. Let me indicate why, by first stating what the consequences of applying Bayes nets algorithms to a system that does not satisfy Bayes nets assumptions are. From a theoretical point of view, it must be clear that Bayes nets algorithms used under those circumstances will not output an adequate representation of the real causal structure. Besides, the absence of beforehand information concerning violations of Bayes nets assumptions makes it impossible to draw an upper limit to the deviation of the output from the true causal structure (under any plausible measure of this deviation). Important practical consequences follow. Indeed, if A causes B, acting on A is a good way to affect B.[15] Then wrongly believing that A causes B can lead to spoil time, energy and money in trying to affect B by modifying A. The losses are all the more dramatic and all the less affordable since the subject matter comes under the social sciences – which were assumed from the beginning in the present paper. To make it short, consequences of undue application of Bayes nets algorithms must not be overlooked. Now, whatever good reasons you may have to do it, resorting to Bayes nets algorithms each time no violation of Bayes nets assumptions is detected beforehand inevitably leads to apply those algorithms to systems failing to satisfy Bayes nets assumptions. What is more, in the absence of any non causal criterion for the satisfaction of those assumptions, any system to which Bayes nets algorithms are actually applied becomes suspect of being one of those problematic cases. With regard to the just discussed consequences of undue application of Bayes nets algorithms, this suspicion cannot be tolerated. The conclusion is that

[13]The current discussion concerning the relationship between the CMC, manipulability and modularity ([Cart02], [Cart06], [Haus99], [Haus04], [Steel06]) may contribute to the formulation of such a criterion. But it seems clear to me that it has not yet, and I cannot see along which lines such a solution could emerge.

[14]The principal two arguments are as follows: 1) according to Bayes nets proponents, Bayes nets assumptions are satisfied by most systems (references were already given in the introduction of the paper) – and, they would probably argue, by nearly all causally sufficient ones; 2) following [Will02] §4, it can be argued that the CMC must be accepted as a "default assumption" under objective Bayesianism. Both of these arguments call for specific comments, which will not be exposed here.

[15]This is a consequence of the commonly accepted relationship of causality with agency.

one should renounce to Bayes nets causal inference.

This conclusion holds for Bayes nets causal inference as it was described in the preceding section of the paper. Yet it may not extend to all possible uses of Bayes nets algorithms. Indeed, several authors have envisaged to integrate Bayes nets algorithms to a wider methodology for causal inference.[16] The proposition is most carefully defended by Williamson[17], under the following form: use Bayes nets algorithms as a first exploratory step of causal inference, then deduce predictions from their output, test those predictions against evidence, consequently amend the hypothesized structure, and finally return to the deduction step for the thus obtained structure. My discussion of this proposition will focus on those consequences that are drawn from the hypothesized causal model. Williamson identifies three kinds of such consequences, respectively deriving from: "supposed connections between causality and probability"[18], the usual correspondence of causal claims with physical processes linking causes to effects and the "close relationship [causality has] with agency"[19]. Now, the latter two facts are clearly irrelevant when one is interested in testing hypothesized causal relations in the social sciences. On the other hand, "supposed connections between causality and probability" are encapsulated in Bayes nets assumptions. As a consequence they will not be violated by structures just output by Bayes nets algorithms, and testing for them is of no use. What remains, then, is Williamson's later suggestion to remove from the amended structure those arrows representing dependencies that are found to admit of a non-causal explanation. I consider that this will not be enough to deal with all errors resulting from undetected violations of Bayes nets assumptions – and to reach a correct answer. As a consequence, I contend that the difficulty raised by the absence of a non causal criterion for the satisfaction of Bayes nets assumptions is not overcome by propositions of the same kind as Williamson's one.

Does it follow that Bayes nets algorithms must be altogether abandoned by anyone who takes seriously our current incapacity to identify systems satisfying Bayes nets assumptions? I would like to show that it does not. More precisely, I will indicate a way in which Bayes nets algorithms can contribute to inference of causal knowledge from observational statistical data in spite of the difficulty with Bayes nets assumptions. It is clear from what precedes that this implies that Bayes nets algorithms are integrated to a wider causal search procedure – like Williamson's – and that this procedure is such that Bayes nets algorithms are run only when Bayes nets assumptions

[16] This answer is developed in particular by [Glym88]pp. 428-429 and in [Will02].
[17] [Will02] section 3.
[18] [Will02]p. 7.
[19] [Will02]p. 8.

are satisfied – unlike Williamson's. In other words, one has to envisage a mixed methodology such that a causal model is produced first and Bayes nets algorithms are run, if possible, subsequently. The most natural idea concerning the production of this initial causal model is to resort to the path analytic methods that have been discussed in section 2. of the present paper. This leads to propose the following methodology: given a system **S** represented by the causally sufficient set of variables **V**,

1". perform steps 1'. to 9'. of path analytical causal inference methodology. Let **M** be the model that is output by this procedure;

2". test whether Bayes nets assumptions would hold for **S** in case it would be correctly represented by **M**. If the test is not passed, then accept **M** as your causal model; if it is, then go to 3".;

3". perform steps 2. to 4. of Bayes nets causal inference for **V**.

- If the output of this procedure is not compatible with **M** – that is if **M** does not belong to the set of directed acyclic graphs represented by the output pattern –, then consider **M** as refuted. Therefore, come back to 9'. in order to produce an alternative causal model. If a model equivalent to **M** seems satisfactory, then come back to 1". with this model; if not, re-iterate steps 2'. to 9'. with new specifications in 1'. and come back to 1". with the new path analytic output;
- If the output of the procedure is compatible with **M**, then accept **M** as your causal model.

The proposed procedure is clearly hypothetico-deductive. Path analytic methods enable to hypothesize a causal model **M**. If it cannot refuted that Bayes nets assumptions hold in case **M** is correct, the hypothesis that it is indeed correct leads to predict that **M** is compatible with the output **P** of Bayes nets algorithms is a consequence[20]. Then Bayes nets algorithms serve in the testing part of the procedure, with recommendations in 3". following: **M** should be rejected when incompatible with **P** and is further corroborated when compatible with **P**. Now the hypothetico-deductiveness of the procedure sounds problematic in as much as all the nice features of Bayes nets causal inference that were highlighted in section 2. precisely stemmed from the deductiveness of the procedure. Actually, those features are lost in the context of the present proposition:

[20]The consequence is not a logical one since the test for the validity of Bayes nets assumptions is open to statistical errors.

- the procedure is not data-based. Rather, it depends on the models the scientist has been able to envisage in the same way as path analytic causal inference does;

- although Bayes nets algorithms still output a pattern, the whole methodology outputs a single model. This point must be explained. Imagine that the output **M** of path analytic causal inference is such that Bayes nets assumptions would be satisfied if **M** were correct. Suppose further that **M** is compatible with the pattern output of Bayes nets algorithms. Then one may wonder why all the models compatible with this pattern cannot be taken as serious candidate models. I see two reasons why they cannot be. First, if plausible, those models have good chance to have been taken into account 1". If it is actually the case, I cannot see why they should be discussed once again after Bayes nets algorithms have been run. Second, even in case one of those models is correct, then nothing guarantees that Bayes nets assumptions are satisfied – contrary to what happens if **M** is correct. As a consequence, compatibility with the output of Bayes nets algorithms does not provide any of those models with any kind of inductive support;

- as a direct consequence of its being hypothetico-deductive, the proposed methodology does not output a model which must be taken as true as soon as the identified zero correlations are. The output model is only well-corroborated by the statistical analysis of data.

The scene, then, is not very heartening. For one thing, the use I envisage for Bayes nets algorithms fails to retain the interesting features of Bayes nets causal inference as applied to systems satisfying Bayes nets assumptions. For another thing, they do not perform causal inference properly speaking anymore. What they do is providing with an additional test for some (but not all) of the models already output by model-based usual path analytic methods – a contribution which obviously is not up to the initial ambitions. Yet the proposed methodology also has non negligible assets. Primarily, it settles the problem of identifying systems that satisfy Bayes nets assumptions – a problem whose important negative consequences we saw to be commonly overlooked. Actually, Bayes nets algorithms are run only when one has good reasons to believe that Bayes nets assumptions are satisfied, more precisely when it cannot be refuted that the assumptions hold in case the model output by path analytic causal inference is correct. Positively, Bayes nets algorithms actually contribute to causal inference when used in the proposed way: path analytic causal inference being a non-deductive procedure, its results always

call for more testing and any kind of further evidence in their favour is welcome and useful. Furthermore, there is no reason to believe that the test as such should be a trivial one. Quite the opposite, chancy compatibility of path analytic and Bayes nets outputs[21] is quite improbable since 1) path analytic causal inference relies on principles differing from Bayes nets ones and 2) Bayes nets algorithms enable to rule out a significant number of candidate causal models. Accordingly, the test relying on Bayes nets algorithms should enable to reject a certain number of models while providing others with substantial inductive support. This contribution is available each time Bayes nets assumptions are satisfied – in most cases according to one of the most common arguments of Bayes nets proponents.

A.1.4 Conclusion

In the present paper I have tried to assess how Bayes nets algorithms can contribute to causal inference. Taking for granted that Bayes nets assumptions sometimes hold and sometimes do not hold led to a two-pronged approach of the question. In section 2., I focused on those systems that satisfy Bayes nets assumptions. This enabled to pinpoint the exact ways in which Bayes nets causal inference is superior to current social sciences procedures. A methodological rationale was given for this superiority. Then, in section 3., I came to the general case through the analysis of the impact of the previous restriction. The difficulty quickly appeared to consist not only of the non universal validity of Bayes nets assumptions, but also of our present incapacity to determine beforehand whether a given system satisfies them or not. The difficulty and existing propositions aiming at overcome it were analyzed. This led me to introduce a mixed methodology for causal inference in which Bayes nets algorithms are run only after good reasons have been provided for the validity of Bayes nets assumptions. Finally, it was explained how Bayes nets algorithms contribute to causal inference in this context.

The present analysis obviously suffers from restriction to the social sciences, and to cases without any latent variables. These restrictions were formulated as a consequence of the impossibility to give in the present paper an extensive treatment of the problem under consideration. Now lifting them, as well as extending the analysis to algorithms learning Bayesian networks through Bayesian techniques, should form the subject matter of subsequent work.

[21] By "chancy compatibility", I refer to a compatibility that would not stem from those outputs being correct.

A.2 Is determinism more favourable than indeterminism for the causal Markov condition?

Abstract

The present essay comments on a paper by Daniel Steel published in the January 2005 issue of BJPS. In this paper, Steel claims to generalize an usual result from the deterministic to the general case by proving that *any* system with jointly independent exogenous variables satisfies the Causal Markov Condition.

It is my contention that the result Steel claims to prove is false unless one is prepared to abandon standard causal modelling terminology. Correlatively, I argue that the most fruitful aspect of Steel's article consists of a realist conception of error terms and I show how this conception sheds new light on the relationship between determinism and the Causal Markov Condition.

A.2.1 Introduction

Several authors have recently suggested that Bayesian networks could significantly help solve the classic problem of inferring causal knowledge from statistical data.[22] But Bayesian networks can be used for causal inference purposes only when the Causal Markov Condition ("CMC" from now onwards) holds. To make it brief, this condition states that every phenomenon is probabilistically independent from all its non-effects conditional on its direct causes. Its truth remains a much debated question.

In this debate, a traditional argument of CMC proponents is that the CMC is true for systems which are deterministic and whose exogenous variables are jointly independent.[23] By contrast, CMC opponents often rely on indeterministic counterexamples. Now, in "Indeterminism and the Causal Markov Condition"[24], Daniel Steel claims to establish that the determinism clause of the case for the CMC is unnecessary – that the CMC is true as soon as exogenous variables are jointly independent. If correct, Steel's thesis should constitute a very important contribution to the CMC debate, and to the larger debate on the possibility of inferring causal knowledge from statistical data.

[22]Major contributions are: [Pearl 2000] and [Spirtes et al. 1993].

[23]The claim is even broader than that, since it also extends to "pseudo-indeterministic" systems. Yet these cases will not be discussed in the present paper.

[24][Steel 2005].

In the present essay, I ponder the correctness of Steel's claim and the contribution of Steel's paper to the CMC debate. Section A.2.2 is a presentation of the CMC-determinism issue independently from Steel's article. Then, Section A.2.3 is a review of Steel's argument. Section A.2.4 assesses it, and leads to the conclusion that it cannot be accepted – at least if standard terminology is not to be abandoned. Nevertheless, Section A.2.5 supports that Steel makes an interesting suggestion as to the representation of causal systems, and shows how this suggestion sheds new light on the relationship between determinism and the CMC.

A.2.2 Determinism and the CMC

The CMC is the causal version of the more general Markov Condition. Therefore I start with a definition of the latter. This is a property of ordered pairs composed of 1) a directed acyclic graph and 2) a probability distribution, both defined over the same set of variables:

Definition 1 (Markov Condition) *Let $\boldsymbol{V}$ be a finite set of discrete variables*[25]*, G a directed acyclic over $\boldsymbol{V}$ and p a probability distribution over $\boldsymbol{V}$.*
(G, p) satisfies the Markov Condition if and only if every variable in $\boldsymbol{V}$ is probabilistically independent from all its non-descendants in G conditional on its direct parents in G.

Let us now consider a "system" – that is, in the present context, any group of causally interrelated phenomena. A simple example consists of a forest that can catch fire because of either lightning, or a match lit by an arsonist.[26] This system can be represented by:

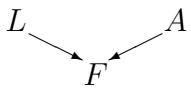
1. the variables L, A, F respectively representing whether lightning strikes, whether an arsonist lights a match in the forest, and whether there is a forest fire[27]. More precisely, L is the binary variable which takes the value 1 if and only if lightning strikes and 0 otherwise, and so on for A and F;

[25] All sets of variables that are considered afterwards are *finite* sets of *discrete* variables. The restriction to discrete variables is not fundamental andis adopted only for simplicity's sake.

[26] This example is given by Halpern and Pearl. See [Halpern and Pearl 2005] p.848.

[27] These variables are meant to represent relevant observable aspects of the situation under consideration. Note that other variables could be considered, giving rise to different representations of one and the same system. This question will not be tackled in the present paper, but I will rather assume that the set of variables representing relevant observable aspects of a given system is univocally determined.

2. the directed graph



   representing the direct causal relationships amongst the variables L, A and F[28]. This graph is "the causal graph" over $\{L, A, F\}$ – or, equivalently, the causal graph for the system under consideration. In the present case, it is acyclic;

3. the probability distribution over $\{L, A, F\}$[29].

More generally, any system can be represented by a triple composed of a set of variables and of the causal graph and probability distribution over this set. As a consequence, the causal version of the Markov Condition can be defined as a property of systems:

Definition 2 (Causal Markov Condition) *Let S be a system represented by the set of variables $\boldsymbol{V}$ and the acyclic causal graph CG and probability distribution p over $\boldsymbol{V}$.*
Then S satisfies the Causal Markov Condition if and only if (CG, p) satisfies the Markov Condition.

A few definitions are necessary in order to set out the usual result concerning the CMC and determinism. Let S be a system, whose relevant observable aspects are represented by the variables in the set $\mathbf{V}$. Then:

- S is *acyclic* if and only if G is acyclic;

- S is *deterministic* if the values of the $\mathbf{V}$ variables having direct causes in $\mathbf{V}$ are functionally determined by the values of their direct causes in $\mathbf{V}$;

- the *exogenous* variables of S are those variables of $\mathbf{V}$ that do not have any direct cause in $\mathbf{V}$;

- the variables of a set are *jointly independent* if any two non-empty distinct subsets of this set of variables are probabilistically independent.

[28]To be precise, one should not talk about causal relations amongst variables, but about causal relations amongst the phenomena those variables represent. Yet, for simplicity's sake, I will allow myself to use the terminology of causality amongst variables, whose real meaning is straightforward.

[29]How this probability distribution should be interpreted will not be discussed here.

Now the classic result to which Steel refers to and beyond which he pretends to go can be stated as follows:

Theorem 1 (Classic result) *Acyclic deterministic systems with jointly independent exogenous variables satisfy the CMC.*

Here is a proof for Theorem 1:

> Let S be an acyclic deterministic system represented by $\{\mathbf{V}, G, p\}$. Let us assume that the exogenous variables in $\mathbf{V}$ are jointly independent. S satisfies the CMC if and only if any variable in $\mathbf{V}$ is independent for p of its non-descendants in G conditional on its direct parents in G. Then, let us consider any variable V_1 in $\mathbf{V}$ and show that it is indeed independent of its non-descendants in G conditional on its direct parents in G.
>
> - if V_1 is endogenous, then its value is functionally determined by the values of its direct parents in G. Therefore it is independent of any of its non-descendants in G when one conditionalizes on the set of its direct parents in G;
> - if V_1 is exogenous, the set of its direct parents in G is empty. Therefore one must simply show that V is (unconditionally) independent of any of its non-descendants in G.
>   Let V_2 be such a variable.
>   - if V_2 is also exogenous, the independence of V_1 and V_2 stems from the hypothesis of joint independence of exogenous variables;
>   - if V_2 is endogenous, by the determinism hypothesis, its value is determined by the value of the set of its direct causes in $\mathbf{V}$. Beyond, and by acyclicity, this value is determined by that of a set of exogenous variables of S, say $\mathbf{W}$. V_1 does not belong to $\mathbf{W}$ since V_2 is not a descendant of V_1. Then the joint independence hypothesis entails that V_1 is independent from $\mathbf{W}$ and therefore from V_2.
>
> As a consequence, any variable in $\mathbf{V}$ is independent of its non-descendants in G conditional on its direct parents in G: S satisfies the CMC.

I shall now turn to Steel's paper.

A.2.3 Steel's argument

What Daniel Steel claims in "Indeterminism and the Causal Markov Condition" is that the determinism clause in Theorem 1 is superfluous. In other words, a system would satisfy the CMC as soon as its exogenous variables are jointly independent. The present section consists of a presentation of the argument Steel has in favour of this claim.

Steel's argument is stated in the framework of "causal functional models" (CFMs for short). Let me explain what they are by first introducing "functional models" (FMs). A FM is defined over a couple of sets of variables. Let $\mathbf{X}$ be $\{X_1, X_2, \ldots, X_n\}$ and $\mathbf{U}$ be $\{U_1, U_2, \ldots, U_k\}$. Then a functional model over $(\mathbf{X}, \mathbf{U})$ is a pair $(\mathbf{E}, p)$ where:

- $\mathbf{E}$ is a set of n equations such that each X_i appears as a function f_i of a non-empty subset of $((\mathbf{X} \cup \mathbf{U}) - \{X_i\})$;
- p is a probability distribution over $\mathbf{U}$.

An example of an FM over $(\{X_1, X_2, X_3\}, \{U_1, U_2\})$ is composed by the equations:

$$X_1 = f_1(U_1, U_2)$$
$$X_2 = f_2(X_1, U_2)$$
$$X_3 = f_3(X_2)$$

together with a probability distribution p over $\{U_1, U_2\}$.[30] Now $(\mathbf{E}, p)$ over $(\mathbf{X}, \mathbf{U})$ is "causal" if 1) all the equations in $\mathbf{E}$ are causal generalizations and 2) any variable X_i in $\mathbf{X}$ is such that the set $\mathbf{DirectCauses}(X_i)$ of its direct causes in $\mathbf{X} \cup \mathbf{U}$ is included in (but not necessarily equal to) the set of its functional parents in $\mathbf{X} \cup \mathbf{U}$[31].

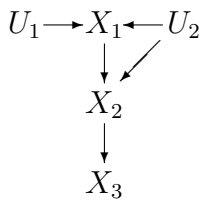
Three remarks must be made before I can actually come to the result established by Steel:

- as noticed by Steel, there is a "straightforward correspondence between functional models and directed graphs"[32]: for an FM $M = (\mathbf{E}, p)$ over $(\mathbf{X}, \mathbf{U})$, there is a unique directed graph G_M over $\mathbf{X} \cup \mathbf{U}$ such that graphical parents in G_M exactly correspond to functional parents in $\mathbf{E}$. For our example FM, the corresponding directed graph is as follows:

[30] This example makes clear that there is no one-to-one correspondence between $\mathbf{X}$ and $\mathbf{U}$ variables of a given FM. This is an important difference between Steel's functional models and the more traditional causal models which will be considered later on.

[31] [Steel 2005] p.9.

[32] [Steel 2005] p.7.



If G_M is acyclic (as is the case here), M as well as systems represented by M will themselves by labeled "acyclic";

- if M is a *causal* functional model, then G_M is an over-graph of the causal graph CG_M over $\mathbf{DC}_M = \bigcup_{X \in \mathbf{X}} \mathbf{DirectCauses}(X)$. As will be clearer soon, CG_M is the causal graph for systems represented by M;

- if G_M is acyclic, then the equations in $\mathbf{E}$ can be restated in such a way that $\mathbf{X}$ variables are functions only of the $\mathbf{U}$ variables from which they descend[33]. As a consequence, p univocally extends to a probability distribution $p\prime$ over $\mathbf{U} \cup \mathbf{X}$, and therefore to a probability distribution $p\prime\prime$ over $\mathbf{DC}_M$.

In terms of CFMs and following the notations already introduced, the result established by Steel is as follows:

Theorem 2 (Steel's result) *Let $M = (\boldsymbol{E}, p)$ be a CFM over $(\boldsymbol{X}, \boldsymbol{U})$.*
If G_M is acyclic and variables in $\boldsymbol{U}$ are jointly independent for p, then $(CG_M, p\prime\prime)$ satisfies the Markov Condition.

In terms of systems[34], it can be formulated as:

Theorem 3 *Let S be a system represented by the CFM $M = (\boldsymbol{E}, p)$ over $(\boldsymbol{X}, \boldsymbol{U})$.*
If S is acyclic and the variables in $\boldsymbol{U}$ are jointly independent for p, then S satisfies the CMC.

The proof given by Steel for his result is not completely convincing to me. Yet the result can also be established along the lines indicated by Pearl for a slightly different result.[35] Moreover, the proof thus obtained makes particularly clear what the rationale for Steel's result is. This proof is made up of:

[33] On this point, see: [Steel 2005] p.8.

[34] Steel does not make the distinction between systems and models representing them. This is unproblematic as long as one assumes that the correspondence between systems and models is functional, which he implicitly does. Yet for reasons that will become clear later on, I will keep systems distinct from models.

[35] [Pearl 2000] p.30.

1. a proof of the fact that $(G_M, p\prime)$ satisfies the Markov Condition – in other words: a proof of the fact that $(G_M, p\prime)$ is a Bayesian network. This first proof runs exactly as the proof that was given for Theorem 1 but with the role of the determinism clause played by functional determination. This last point should be noticed, since it will reveal essential in the sequel of the paper;

2. a derivation of the probabilistic conditional independencies that are required in order for $(CG_M, p\prime\prime)$ to satisfy the Markov Condition. Those independencies are established by d-separation in the Bayesian network $(G_M, p\prime)$.

Now what is the relationship between Steel's result and the determinism / indeterminism issue concerning the CMC? The answer consists of the following proposition: CFMs can represent indeterministic as well as deterministic systems. In order to make it clear, Steel gives the following example:

> Imagine a special type of car, the quantum car. The ignition of the quantum car works by means of a fundamentally indeterministic process: when the key is turned, there is an irreducible probability of .85 that the car will start.[36]

Then Steel explains that the system constituted by the "quantum car" can be represented by a CFM M over $(\{X_1, X_2\}, \{U_1, U_2\})$, where

- X_1 is "a binary variable indicating whether the key is turned ($X_1 = 1$ indicates that it has been)"[37];

- X_2 is "a binary variable representing whether the car starts ($X_2 = 1$ indicates that it does)"[38];

- (as far as I understand Steel's treatment of the example) U_1 represents the reasons why the car may be started;

- U_2 is a binary variable whose possible values are 0 and 1 and which represents whether the car starts once the key has been turned.

M is composed of the equations

$$X_1 = f_1(U_1)$$
$$X_2 = U_2.X_1$$

[36] [Steel 2005] p.13.
[37] [Steel 2005] p.13.
[38] [Steel 2005] p.13.

together with a probability distribution p over $\{U_1, U_2\}$ which is such that $p(U_2 = 1) = .85$. With p defined in this way, it becomes clear that U_2 represents the probabilistic nature of the action of X_1 on X_2. More generally, the example makes clear how any indeterministic system can be represented by a causal functional model, with **U** variables representing the probabilistic nature of the action of indeterministic causes on their effects. As a consequence, Steel considers that Theorem 2 implies that the CMC is true for any acyclic system with jointly independent exogenous variables, be it deterministic or not.

A.2.4 Assessment of Steel's argument

Obviously, a necessary condition for Steel to have actually proved that any acyclic system with jointly independent exogenous variables satisfies the CMC is that the **U** variables of a CFM are the exogenous variables of systems represented by that CFM. This is assumed by Steel. Yet he has an hesitation when introducing them: he talks about "a set of *exogenous variables* or error terms"[39]. It seems to me that this hesitation is very meaningful. Yet this can be explained only after I have told the standard story about exogenous variables, error terms, and the way they differ. The standard definition of exogenous variables has been introduced in Section A.2.2. One has 0) a system S, 1) a set $\mathbf{V} = \{V_1, V_2, \ldots V_n\}$ of variables representing its relevant observable aspects, 2) a graph CG representing the direct causal relationship amongst the variables of **V**. Then the exogenous variables of S are those variables in **V** that do not have any parents in CG. All this has already been stated. But from there on one can go one step further in the representation of S and consider a set of n equations such that each V_i constitutes the right-hand-side of exactly one equation and is a function of its direct parents in CG plus one variable T_i. This set of equations together with the probability distribution over **T** constitutes a representation of the system under consideration which is common in the field of causal modelling[40]. Therefore this pattern of representation will be referred to as that of "standard causal functional models" ("standard CFMs" for short). At least to begin with, standard CFMs should be carefully kept distinct from Steel's CFMs[41]. The **T** variables of standard CFMs have a precise function: they enable a functional representation of the relations between **V** variables. Indeed, each T_i represents everything that contributes to the determination of the value of V_i and yet is not represented by the variables in $\mathbf{V} - \{V_i\}$. More

[39] [Steel 2005] p.5.

[40] See for instance Pearl's "causal models" as defined in [Pearl 2000] p.27.

[41] That is the reason why I have adopted the notation T_I instead of the U_i usual one.

precisely, and following Cartwright's analysis, **T** variables "represent, in one heap, measurement errors, omitted factors, and whatever truly probabilistic element there may be in the determination of an effect by its causes"[42]. **T** variables are those usually called "error terms".

This was the standard story. Let us now come back to Steel's one. Are his **U** variables "*exogenous variables* or error terms"[43]? Stated in other words: what do Steel's **U** variables represent? It is clear from Steel's paper that **U** variables can represent two very different kinds of things:

- some of them, like U_1 in the quantum car example, represent relevant observable aspects of the system under consideration that happen not to have direct causes in the system. Those variables are exogenous variables in the standard sense of the term;

- the other ones, like U_2, represent the way probabilistic causes act on their effects. Those variables are not exogenous variables since they do not represent observable relevant aspects of the system under consideration. Moreover, they represent one of the three kinds of influences that error terms in the standard sense represent. But they represent *only one* of the three kinds of influences that are represented "in one heap" by error terms. Failing to represent everything that error terms usually represent, **U** variables of this second kind are not error terms standardly speaking.

This analysis easily accounts for the hesitation of Steel when introducing his **U** variables: **U** is composite object whose elements either are exogenous variables in the standard sense, or represent part of what standard error terms represent. As a consequence, they cannot be called "exogenous variables" if one is conforming to standard terminology. Stated in another way, Steel's result actually amounts to the truth of the CMC for *any* acyclic system with jointly independent exogenous variables only if one abandons the standard pattern of functional representation of systems and the sense "exogenous variables" has in this context. Therefore it cannot be said without qualification that Steel has established the superfluity of the determinism clause in Theorem 1. Omitting qualification, as Steel does, is seriously misleading as to what the result actually is.

It could be that all this does not matter much. This would be the case if Steel showed the way towards a proof of the result he claims to have established. More likely, it may be possible to prove that the CMC is true

[42] [Cartwright 1999] p.9.
[43] [Steel 2005] p.5.

for any system with jointly independent variables in a way similar to the one I proved Steel's result. This seems all the more likely since the sketch of proof given by Pearl is modular and has already proven fecund enough in producing proofs for Pearl's and Steel's results, which are quite different. Yet a quick examination of the sketch reveals that one can obtain a final result in terms of joint independence of *exogenous variables* (standardly defined) only if one adopts a pattern of representation from which error terms are absent. But we saw that the first part of the proof relies in an essential way on the functional determination of the values of the effects by those of their direct causes. In the absence of error terms, such a functional determination is exactly equivalent to determinism. In other words, Pearl's sketch of proof is no use for one who wants to establish that *all* acyclic systems with jointly independent exogenous variables satisfy the CMC. Of course this does not imply that the claim is false. What implies it by contrast is the fact that, under standard terminology, it remains possible for an acyclic system with jointly independent exogenous variables to fail to satisfy the CMC – provided it is indeterministic. This is the case of Nancy Cartwright's classic example:

> Cheap-but-Dirty employs a genuinely probabilistic process to produce the chemical [a chemical that is consumed in a given sewage plant]. The probability of getting the desired chemical on any day the factory operates is eighty percent. [...] [Moreover] pollutants are emitted as a by-product whenever the chemical is produced.[44]

Relevant observable aspects of the systems are represented by three binary variables with value 0 or 1: O indicating whether Cheap-but-Dirty operates, S indicating whether sewage is produced, and P indicating whether pollutant is produced. It is easily seen that only one of them is exogenous in the standard sense: O. Hence, trivially, exogenous variables are jointly independent. Moreover, the system is clearly acyclic. And yet the system does not satisfy the CMC since O does not screen off P from S:
$p(P = 1, S = 1|O = 1) = .8$, whereas
$p(P = 1|O = 1) \times p(S = 1|O = 1) = .8 \times .8 = .64$.
As a consequence of this counterexample, the result Steel claims to have established remains false under standard terminology.

Acyclicity and joint independence of exogenous variables are sufficient for the CMC only if one accepts as exogenous variables, variables that represent the way probabilistic causes act on their effects. It could be argued that there is nothing wrong with this and, quite the opposite, that this modelling proposition constitutes the very innovation in Steel's paper. This is precisely

[44] [Cartwright 1999] p.7.

the position adopted by Steel himself: "The basic insight is that exogenous variables in an FM can be interpreted either as representing causes or genuine indeterminism"[45]. The matter is that variables representing the way probabilistic causes act on their effects already existed before Steel's paper, and that they were never called "exogenous variables"; conversely, "exogenous variables" and "systems with jointly independent exogenous variables" already had a meaning before Steel's paper, and this meaning is different from the one they have in Steel's paper. Moreover, Steel does not give any independent justification for substituting his interpretation of exogenous variables to the usual one. As a consequence, his modelling innovation should not be accepted unless the associated terminology were carefully distinguished from the usual one and the author were careful not to claim to enter a pre-existing debate – two things that Daniel Steel fails to do.

A.2.5 What Steel's paper suggests

I would like to end with more positive considerations. Indeed I think that Steel's paper actually contributes to the CMC debate, in spite of the difficulties hitherto highlighted. As the difficulties, the contribution lies in Steel's **U** variables – and to be more precise, in those **U** variables that are neither exogenous variables nor error terms in the standard sense of these terms. I have already stated that those variables represent only one of the three kinds of influences that standard error terms represent "in one heap", and this appeared to be problematic. But there is another way of looking at this: while standard error terms clearly fail to represent any real entity, Steel's **U** variables lean towards realism in the use of error terms – by which is meant that error terms are used in such a way that the structure of a causal model matches the real causal structure it represents. More specifically, Steel's **U** variables suggest to disjoin the influences that are represented "in one heap" by standard error terms, that is to represent distinct influences by distinct error terms. Following this suggestion, there would correspond to each variable V representing a relevant observable aspect of the system under consideration:

- an error term representing the forgotten direct causes of what V represents;

- an error term representing the possible errors in the measurement of V;

[45] [Steel 2005] p.4.

- for each of V's probabilistic direct causes, an error term representing the way it acts on V.

These non-standard error terms lead to define non-standard causal functional models. Given a system S whose relevant observable aspects are represented by $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$, a non-standard CFM for S is $(\mathbf{E}, p)$ where:

- $\mathbf{E}$ is a set of n equations such that each V_i is the right-hand-side of exactly one equation and appears as a function of 1) its direct causes in $\mathbf{V}$ and 2) non-standard error terms $T_{i,1}, T_{i,2}, \ldots, T_{i,m}$. Observe that the number m of error terms depends on which variable V_i is considered and can therefore be noted "$\varphi(i)$";

- p is the probability distribution over the set $\mathbf{T}$ of non-standard error terms.

Non-standard CFMs complete Steel's CFMs so as to ensure that everything standard CFMs represent is actually represented. This is enough for non-standard CFMs not implying the important redefinition of usual terminology that Steel's CFMs did convey, and that was identified as problematic. But as for the rest, the two patterns of representation are largely similar. First and foremost, non-standard CFMs allows a characterization of determinism that relies on the very intuition that underlain Steel's "Definition of Deterministic Functional Models"[46]:

Definition 3 (Characterization of Determinism) *A system is deterministic if it is represented by a non-standard causal functional model with no error term representing the way a probabilistic cause acts on one of its effects.*

Then, as Steel's CFMs, non-standard CFMs are in "straightforward correspondence"[47] with directed graphs. Accordingly, they and the systems they represent will also be labelled "acyclic" when the corresponding directed graph is acyclic. Moreover, acyclicity of the non-standard CFM $M = (\mathbf{E}, p)$ over $(\mathbf{V}, \mathbf{T})$ remains sufficient for the probability distribution p over $\mathbf{T}$ to univocally extend to a probability distribution $p\prime$ over $\mathbf{T} \cup \mathbf{V}$. To finish with, and still as with Steel's CFMs, the graph G_M corresponding to M is an overgraph of the causal graph for systems represented by M. Notice that this causal graph is now univocally determined by M (as the directed graph representing the direct causal relations amongst $\mathbf{V}$ variables that are depicted by $\mathbf{E}$); therefore it can and will be noted: "CG_M".

[46] [Steel 2005] p.9.
[47] [Steel 2005] p.7.

Does this tell us anything new about the relationship between determinism and the CMC? A first step towards an answer to this question consists of the fact that, under the notations that have just been introduced, the following result holds:

Theorem 4 *Let $M = (\boldsymbol{E}, p)$ be an acyclic non-standard CFM over $(\boldsymbol{V}, \boldsymbol{T})$ and p'' the restriction of p' to $\boldsymbol{V}$.*
If variables in $\boldsymbol{T}$ are jointly independent, then (CG_M, p'') satisfies the Markov Condition.

Theorem 4 is the equivalent, in the framework of non-standard CFMs, of Steel's result. It holds for exactly the same reasons and can also be stated in terms of systems:

Theorem 5 *Let S be a system represented by the non-standard CFM M over $(\boldsymbol{V}, \boldsymbol{T})$.*
If M is acyclic and variables in $\boldsymbol{T}$ are jointly independent, then S satisfies the CMC.

By way of illustration, let us consider a classic (non quantum) car, by which is meant a car that starts each time the key is turned and there is petrol in the tank. Let S be this system. Three variables are needed in order to represent S: V_1 representing whether the key is turned, V_2 representing whether there is petrol in the car and V_3 representing whether the car starts. Let $M = (\mathbf{E}, p)$ be the non-standard CFM representing S. Not assuming linearity, equations in $\mathbf{E}$ are as follows:

$$V_1 = g_1(T_{1,1}, T_{1,2})$$
$$V_2 = g_2(T_{2,1}, T_{2,2})$$
$$V_3 = g_3(Y_1, Y_2, T_{3,1}, T_{3,2})$$

with $T_{1,1}$ representing the omitted causes of V_1, $T_{1,2}$ the possible errors in the measurement of the value of V_1, and correspondingly for $T_{2,1}$, $T_{2,2}$, $T_{3,1}$ and $T_{3,2}$. Given acyclicity of S, Theorem 5 states that a sufficient condition for it to satisfy the CMC is that the variables in $\{T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}, T_{3,1}, T_{3,2}\}$ are jointly independent.

Now, this condition can be refined. Indeed, and as already stressed by Cartwright[48] in another context, what is needed for a proof in the style of Pearl's one to be possible is only the joint independence of the "net effects" of error terms. In our example, "net effects" of error terms correspond to $\{T_{1,1}, T_{1,2}\}$, $\{T_{2,1}, T_{2,2}\}$ and $\{T_{3,1}, T_{3,2}\}$. Then, following Cartwright, it is sufficient that any way of combining them into two non-empty sets is in two *independent* sets of variables. To be explicit, it is sufficient that:

[48] [Cartwright 2001] p.18.

(1) $\{T_{1,1}, T_{1,2}\}$, $\{T_{2,1}, T_{2,2}\}$ and $\{T_{3,1}, T_{3,2}\}$ are pairwise independent;

(2) $\{T_{1,1}, T_{1,2}\}$ is independent from $\{T_{2,1}, T_{2,2}, T_{3,1}, T_{3,2}\}$;

(3) $\{T_{2,1}, T_{2,2}\}$ is independent from $\{T_{1,1}; T_{1,2}, T_{3,1}, T_{3,2}\}$ and

(4) $\{T_{3,1}, T_{3,2}\}$ is independent from $\{T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}\}$.

Joint independence of $\{T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}, T_{3,1}, T_{3,2}\}$ is too strong a condition; in particular it is not necessary for $T_{1,1}$ to be independent from $T_{1,2}$, for $T_{2,1}$ to be independent from $T_{2,2}$, or for $T_{3,1}$ to be independent from $T_{3,2}$. Stated in terms of any system represented by a non-standard CFM over $(\mathbf{V}, \mathbf{T})$, a proof in the style of Pearl's one only requires the following: for any two distinct non-empty $\mathbf{W}, \mathbf{W}\prime \subset \mathbf{V}$, the sets of error terms respectively corresponding to the variables in $\mathbf{W}$ and to the variables in $\mathbf{W}\prime$ are independent. Rigorously:

Theorem 6 *Let $M = (\boldsymbol{E}, p)$ be an acyclic non-standard CFM over $(\boldsymbol{V}, \boldsymbol{T})$, with $\boldsymbol{V} = \{V_1, V_2, \ldots, V_n\}$.*
If for any distinct i, j, k, l in $[|1, n|]$, $\{T_{i,1}, \ldots, T_{i,\varphi(i)}, \ldots, T_{j,1}, \ldots, T_{j,\varphi(j)}\}$ is independent from $\{T_{k,1}, \ldots, T_{k,\varphi(k)}, \ldots, T_{l,1}, \ldots, T_{l,\varphi(l)}\}$,
then $(CG_M, p\prime\prime)$ satisfies the Markov Condition.

For simplicity's sake, let me refer to the antecedent of the implication stated by Theorem 6 as: "M satisfies **JI**". Now, in terms of systems, we have:

Theorem 7 *Let S be a system represented by the non-standard CFM M. If M is acyclic and satisfies **JI**, then S satisfies the CMC.*

It is my contention that, under Definition 3, Theorem 7 reveals a sense in which determinism is more favourable than indeterminism to the CMC. More specifically, while I agree with Steel writing that

> It is hard to see why deterministic causal systems would be more likely acyclic than indeterministic ones[49],

I will explain why deterministic systems are more likely than indeterministic ones to have non-standard CFMs satisfying **JI**. To that effect, let me introduce a system $S\prime$ which is identical to the previously introduced S except for the fact that turning the key has "an irreducible probability of .85" of doing its electrical job (exactly as in Steel's quantum car). $S\prime$ is represented by a non-standard CFM $M\prime = (\mathbf{E}\prime, p\prime)$. Under previous notations, equations in $\mathbf{E}\prime$ are as follows:

[49] [Steel 2005] p.16.

$$V_1 = g_1(T_{1,1}, T_{1,2})$$
$$V_2 = g_2(T_{2,1}, T_{2,2})$$
$$V_3 = g_3(T_{3,3} \times V_1, V_2, T_{3,1}, T_{3,2})$$

with $T_{3,3}$ a binary variable with possible values 0 and 1 and such that $p\prime(T_{3,3} = 1) = .85$. As a result, $M\prime$ satisfies **JI** if:

(1′) $\{T_{1,1}, T_{1,2}\}$, $\{T_{2,1}, T_{2,2}\}$ and $\{T_{3,1}, T_{3,2}, T_{3,3}\}$ are pairwise independent;

(2′) $\{T_{1,1}, T_{1,2}\}$ is independent from $\{T_{2,1}, T_{2,2}, T_{3,1}, T_{3,2}, T_{3,3}\}$;

(3′) $\{T_{2,1}, T_{2,2}\}$ is independent from $\{T_{1,1}; T_{1,2}, T_{3,1}, T_{3,2}, T_{3,3}\}$ and

(4′) $\{T_{3,1}, T_{3,2}, T_{3,3}\}$ is independent from $\{T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}\}$.

Now suppose that all these independencies hold. Then, the independencies stated by (1) to (4) all hold. This stems from the simple probabilistic following fact: for any two sets of variables $\mathbf{Y}$ and $\mathbf{Z}$ and any variable V, if $\mathbf{Y} \cup \{V\}$ is independent from $\mathbf{Z}$, then $\mathbf{Y}$ is independent from $\mathbf{Z}$. In other words, if $M\prime$ satisfies **JI**, then M does too. The converse is not true: $\{T_{1,1}, T_{1,2}\}$ (for instance) can be independent from $\{T_{3,1}, T_{3,2}\}$ while not being independent from $\{T_{3,1}, T_{3,2}, T_{3,3}\}$. These are no facts particular to our car example. Indeed, the example makes clear that for any two S and $S\prime$ differing only by the probabilistic nature of some causal relations in $S\prime$, the non-standard CFMs M and $M\prime$ representing them only differ by the absence from M of the error terms standing for indeterminism in $M\prime$. Note that the non-standard error terms representing indeterminism in $S\prime$ never *substitute for* error terms present in the deterministic case; they can only *add to* them. The consequence is that M satisfies **JI** if $M\prime$ does while the converse does not hold. In this sense, non-standard CFMs representing deterministic systems are more likely than non-standard CFMs representing indeterministic systems to satisfy **JI**. To this exact – and acknowledgedly narrow – extent, determinism can be said to be more favourable a context than indeterminism for the truth of the CMC.

A.2.6 Conclusion

Steel's claim to enlarge the usual result concerning the relationship between determinism and the CMC from the deterministic to the general case revealed unacceptable. More precisely, it revealed to be unacceptable under standard causal modelling terminology. Indeed, the analysis showed that the truth of Steel's result relies on a highly uncommon use of "exogenous variables". Yet this use of "exogenous variables" suggested that error terms could be

employed in a much more realist way than the standard one. I defined the pattern of representation that stems from this suggestion, and explained how the determinism-CMC debate can take advantage of this new framework. Determinism was easily characterized under this framework and a sense in which determinism is more favourable a context than indeterminism for the CMC appeared: the sufficient condition we have for the truth of the CMC is such that if two systems differ only by determinism, then the deterministic one satisfies the condition whenever the indeterministic one does – and the converse does not hold.

Hinting at "non-standard causal functional models" and at Definition 3 probably constitute the most significant contribution of "Indeterminism and the Causal Markov Condition" to the debate concerning the CMC in general, and its relationship to determinism in particular. But this Definition, as well as the sufficient condition for the truth of the CMC that was derived from it, are valid only for systems represented by causal functional models – "non-standard" ones in the case in point. Hence, this Definition and the condition derived from it are useful exactly in as much as causal functional models can represent real systems, and are interesting exactly in as much as causal functional models can represent *interesting* real systems. Whether they can is a huge and difficult problem.

A.3 Can there be a propensity interpretation of conditional probabilities?

Abstract

The present paper deals with Humphreys' thesis that the propensity theory of probability does not enable to interpret conditional probabilities. Relying on an analysis of what it is to conditionalize, I argue *contra* Humphreys that there can be a propensity interpretation of conditionalization. In the end, an interpretation is actually suggested and discussed.

A.3.1 Introduction

The propensity interpretation of probability was introduced by Popper during the late 1950s.[50] Its main appeal was (and still is) to let singular probabilities be physical – that is dependent only on the state of the world. Indeed, the fundamental idea behind the interpretation is that probabilities of singular events depend on the physical system that produces those events. For example, the probability of getting a "six" on next throw of a given die depends on the physical properties of the throwing device – that is, in the present case, mainly the physical properties of the die itself and of the surface on which it is to be thrown. Due to its physical properties, the throwing device has a propensity to produce a "six" on next throw, and the probability of getting a "six" on next throw measures this propensity. In short, probabilities measure propensities to produce singular events. To the extent that those propensities differ from physical system to physical system, probabilities depend on the physical system one considers.

Since its introduction, the propensity interpretation of probability has been confronted with many criticisms. Among those criticisms, the most robust as well as central one is undoubtedly "Humphreys' paradox". According to this criticism, there cannot be a propensity interpretation of conditional probabilities. Two important consequences ensue. In the first place, one has to give up the idea of a correspondence between subjective and physical probabilities that would conform to Lewis' Principal Principle.[51] Indeed, rational degrees of conditional belief are conditional probabilities[52] and physical conditional probabilities should be too if something like the Principal Principle were to be retained. In the second place, one has to give up the idea that

[50] [Popp57] and [Popp59].
[51] [Lewis80].
[52] [Tell73].

probability theory as we know it is the theory of all aleatory events.

Now, the present paper calls Humphreys' conclusion into question, and actually suggests a propensity interpretation of conditional probabilities. The originality of the proposed solution lies less in its content, than in a constructivist approach to the problem. More precisely, the solution is built out of an analysis of what it means to interpret conditional probabilities. Accordingly, the first section of the paper is a presentation of Humphreys' paradox, the second section deals with the very notion of interpreting conditional probabilities and the third section sets out a solution to Humphreys' paradox. Finally, a fourth section discusses the proposed solution.

A.3.2 Humphreys' paradox

As noticed by Humphreys himself, there exists a « variety of versions »[53] of the paradox, some formal and some not. I focus on the formal versions because, Humphreys points out, « for those, a satisfactory solution is required »[54]. More precisely, I first focus on the formal paradox as it stands in Humphreys' seminal paper : [Hump85][55]. Once this is done, I explain how the initial paradox can be generalized, giving rise to other formal versions of the paradox.

Humphreys' paradox (in any of its versions) deals with inverse conditional probabilities, that is with conditional probabilities such that the conditioning event is temporally posterior to the conditioned event. A proper example is to be found in [Hump85]. In this paper, Humphreys considers the physical system described as follows :

> A source of spontaneously emitted photons allows the particles to impinge upon the mirror, but the system is so arranged that not all the photons emitted from the source hit the mirror [...] Let I_{t2} be the event of a photon impinging upon the mirror at time $t2$, and let T_{t3} be the event of a photon being transmitted through the mirror at time $t3$ later than $t2$. Now consider the single-case conditional propensity $Pr_{t1}(.|.)$ where $t1$ is earlier than $t2$.[56]

[53] [Hump04] p. 668.

[54] [Hump04] p. 668.

[55] The paper is seminal to Humphreys' paradox in the sense that it is the first paper Humphreys devotes to his objection. Yet he already had the objection and the objection was already known of philosophers of science by the end of the 1970s. The name "Humphreys' paradox" was introduced in [Fetz81].

[56] [Hump85] p. 561.

In this context, $Pr_{t1}(I_{t2}|T_{t3})$ is an inverse conditional probability[57].

Now one can wonder what the value of this probability is. Humphreys' answer to this question is by principle (CI) :

Principle 1 (CI) *If Pr_{t1} is a probability function given a propensity interpretation, E_{t2} and E_{t3} two singular events such that $t1 < t2 < t3$, then $Pr_{t1}(E_{t2}|E_{t3}) = Pr_{t1}(E_{t2}|\overline{E_{t3}}) = Pr_{t1}(E_{t2})$.*

In Humphreys' example, the probability $Pr_{t1}(I_{t2}|T_{t3})$ of the photon impinging upon the mirror at $t2$ given that it is transmitted at $t3$ is exactly the probability $Pr_{t1}(I_{t2})$ of its impinging upon the mirror at $t2$. Humphreys' justification is as follows : « the propensity for a particle to impinge upon the mirror is unaffected by whether the particle is transmitted or not »[58]. In more general terms, (CI) results from the fact that posterior events do not (at least, normally) influence prior events. Justified as it is by Humphreys, (CI) looks sound. The matter is that it is incompatible with some basic properties of standard conditional probabilities. More precisely, Humphreys derives two contradictions. The premises of the first derivation are : (CI) applied to the photon example, the classic theorem on total probability for binary events, and probability ascriptions that immediately stem from the description of the photon example. The second derivation has the same premises except that Bayes' theorem is substituted for the theorem on total probability. This is Humphreys' paradox.

There is only one strategy out of Humphreys' paradox : rejecting (CI). Positively, one may substitute for it another principle for evaluating inverse conditional probabilities under a propensity interpretation. Apart from (CI), the literature on the propensity interpretation has two such principles to offer :

Principle 2 (ZI) *If Pr_{t1} is a probability function given a propensity interpretation, E_{t2} and E_{t3} two singular events such that $t1 < t2 < t3$, then $Pr_{t1}(E_{t2}|E_{t3}) = 0$*

and :

Principle 3 (FP) *If Pr_{t1} is a probability function given a propensity interpretation, E_{t2} and E_{t3} two singular events such that $t1 < t2 < t3$, then $Pr_{t1}(E_{t2}|E_{t3}) = 0$ or 1.*

[57] Humphreys does not distinguish between propensities and probabilities interpreted as propensities. For clarity's sake, I do.

[58] [Hump85] p. 561.

(ZI) is supported in particular by Fetzer[59] and expresses the fact that the influence of a posterior event on a prior event is null. (FP) is supported in particular by Milne[60] and expresses the fact that, at $t3$, E_{t2} has definitely occurred or failed to occur. Obviously, it should be debated which one of principles (CI), (ZI) and (FP) is adequate for evaluating inverse conditional propensities. However, whatever the conclusion of the debate, Humphreys' paradox remains. More precisely, it is shown in [Hump04] that there exist formal versions of the paradox analogous to the intial one, but dealing with (ZI) for the first one and (FP) for the second one. Humphreys' paradox is thus generalized.

At that point, the situation looks quite desperate. Still, two remarks can be made by one who would like to see the paradox solved. In the first place, the generalization of the paradox in [Hump04] does not imply that no solution can ever be found. Humphreys shows that a version of the paradox corresponds any known alternative to (CI); yet he does not show that the paradox cannot possibly be solved. In the second place, I would like to draw attention to an assumption underlying the whole discussion and accepted by all those who take part in it. The assumption is as follows : the propensity interpretation for absolute probabilities (as we described it in the section A.3.1) analytically contains an interpretation of conditional probabilities. The problem, then, is to rightly identify it – and to uncover how it leads to evaluate inverse conditional probabilities. It is my contention that this assumption cannot be accepted, and that no interpretation of conditional probabilities is analytically contained in an interpretation of absolute probabilities. This can be shown be analyzing the very notion of interpreting conditional probabilities.

A.3.3 Interpreting conditional probabilities

It is known since [Lewis76] that conditional probabilities are not probabilities of conditionals. More explicitly, Lewis shows that, except for trivial cases, there does not exist a connective $\Rightarrow$ such that conditional probabilities $Pr(A|C)$ have the same values as absolute probabilities $Pr(C \Rightarrow A)$.[61] As a consequence, conditionalizing does not amount to substituting a complex expression – in the $C \Rightarrow A$ style – for a simpler one, as an argument of an unchanged probability function. Positively, conditionalizing is changing

[59] [Fetz81].

[60] [Milne86].

[61] [Lewis76] pp. 300–303. More precisions about Lewis' results, and in particular about the distinction between the two results he has, are not needed here.

the probability function itself. Interpreting conditionalization, then, is telling how the the probability function is modified by the conditioning event.

Let me illustrate the idea just stated. The illustration is with the subjectivist theory of probability. It is justified by the fact that this theory provably constitutes an interpretation of both absolute and conditional probabilities.[62] Under the subjectivist interpretation, absolute probabilities measure degrees of rational belief. Specifically, let Pr be the function measuring the degrees of rational belief of individual I under the stock of information K. Then, the subjectivist claims, $Pr(.|A)$ is the function measuring I's degrees of rational belief under information $K \cup \{A\}$. In other words, conditionalizing on A amounts to adding A to I's initial stock of information. This makes two points : first that conditionalization indeed has to be interpreted, second that the interpretation indeed consists in explicating how the probability function is modified by the conditioning element.

The subjectivist illustration allows to go a little further in the analysis what of it is to conditionalize – and, consequently, to interpret conditionalization. Indeed, it is clear from the presentation just given that a probability function interpreted in the subjectivist way depends on the individual I one considers, as well as on I's stock of information K. Since they determine the probability function, these could be called its "determinants". They differ from the arguments of the function which, in the subjectivist case, are propositions. Now it appears that interpreting conditionalization consists of telling how an argument redefines the determinant of the probability function. More rigorously, conditionalization is to be interpreted as a function which associates a new determinant to any pair composed of an initial determinant and an argument. Let us put it formally in the subjectivist case. To that effect, let $\mathbf{I}$ be the sets of individuals, $\mathbf{P}$ the set of propositions, and $\mathbf{K}$ the powerset of $\mathbf{P}$. Then the subjectivist interpretation of conditionalization is by the function c_s defined as follows:

$$\begin{array}{rcccc} c_s & : & (\mathbf{I} \times \mathbf{K}) \times \mathbf{P} & \longrightarrow & \mathbf{I} \times \mathbf{K} \\ & & ((I, K), P) & \longmapsto & (I, K\prime) = (I, K \cup \{P\}). \end{array}$$

This leads to a general analysis of an interpretation of probability as consisting of :

1. an interpretation of absolute probabilities. This must specify in particular :

   (a) what kind of objects arguments of probability functions are ;

[62] This is established in [Rams26] for the absolute case and in [Tell73] for the conditional one.

   (b) what kind of objects determinants of probability functions are ;

2. an interpretation of conditionalization as a function from the cartesian product of the set of arguments and the set of determinants, to the set of determinants.

Armed with this analysis, I come back to the propensity interpretation.

A.3.4 A propensity interpretation of conditional probabilities

It is clear from the presentation given in the introduction that under the propensity interpretation of probability :

1. absolute probabilities measure the propensities of physical systems to realize singular events. Hence,

   (a) arguments of probability functions are elements of the set **E** of singular events ;

   (b) determinants of probability functions are elements of the set **S** of physical systems.

   Consequently,

2. conditionalization is to be interpreted as a function c_p from the cartesian product of the set of physical systems with the set of singular events, to the set of physical systems.

At that point, the question of the propensity interpretation of conditional probabilities becomes the question of defining c_p. In other words, analyzing what it is to interpret conditionalization leads us to state the question of the propensity interpretation of conditional probabilities in a way noticeably different from the way it is usually stated. What is at stake is no more whether the interpretation of conditional probabilities that is presumed analytically contained in the propensity interpretation of absolute probabilities is admissible. It becomes constructing an admissible propensity interpretation of conditionalization.

To that effect, I take into consideration the following property of Bayesian conditionalization : for any probability function Pr and any E such that $Pr(E) \neq 0$, $Pr(E|E) = 1$. This imposes a constraint on the function c_p constituting the propensity interpretation of conditionalization. Specifically, c_p must be have the following property:

Property 1 *For any physical system S and any singular event E such the $Pr_S(E) \neq 0$, $Pr_{c_p,S}(E) = 1$.*

Now this suggests to define c_p in the following way : :

$$\begin{array}{rcccl} c_p & : & \mathbf{S} \times \mathbf{E} & \longrightarrow & \mathbf{S} \\ & & (S, E) & \longmapsto & \text{the system the most similar to } S \text{ among those} \\ & & & & \text{giving probability 1 to } E. \end{array}$$

In words, the proposition is to interpret conditionalization as the function that associates to the initial physical system, the system from which it differs the least among those that give probability 1 to the conditioning event. Thus, conditionalization is interpreted as the minimal move required for the satisfaction of Property 1.

This, it seems to me, counts as an asset of the proposed interpretation. Another asset of the proposition concerns inverse conditional probabilities. Humphreys' paradox, it was explained, runs on the question of their evaluation. Now, the proposed interpretation does not give any reason to evaluate them following one of the principles (CI), (FP) et (ZI) discussed above. In other words, the proposed interpretation does not commit to one of the principles for which Humphreys has established a formal paradox. Even better, it apparently commits to no principle at all for the evaluation of inverse conditional probabilities. If this is indeed the case, then no new formal version of Humphreys' paradox can be set against the just proposed interpretation.

A.3.5 Discussion

Although the proposed interpretation has some assets, it is clear that it also raises objections. Some of them are discussed in this final section. The first objection, presumably, would be to introducing the notion of similarity between systems. As an answer, I will not produce a definition of similarity between physical systems. Rather, I claim that the objection is at least seriously weakened by widening the view. More precisely, it is well-known that the leading approach to counterfactuals is through similarity between possible worlds. Now, it seems to me that accepting similarity between systems is no great deal once similarity between possible worlds has been accepted. Reciprocally, rejecting similarity between systems seems to commit to rejection of similarity between possible worlds – and, along with it, of our best analysis of counterfactuals and of an important analysis of causality. As a consequence I do not think that its resorting to similarity between physical systems invalidates the proposed interpretation.

Another difficulty with the proposed interpretation is, precisely, that nothing guarantees that it is indeed an interpretation of conditionalization.

It was constructed in such a way that it adequately accounts for Property 1. But it could have been constructed referring to another property of Bayesian conditionalization. Worse still, I have not given any reason to think that the proposed interpretation indeed accounts for all the properties of Bayesian conditionalization. Once again, I do not have a concluding answer to the objection. Yet I have two arguments to put forward. First, it may be noticed that Property 1 is fundamental to Bayesian conditionalization, and this is a reason for considering it rather than another property of Bayesian conditionalization.[63] My second argument, now, is comparative. The comparison is no more with alternative interpretations of probability, but with the propensity interpretation of *absolute* probabilities. Actually, there does not exist a conclusive argument to the effect that the propensity theory of absolute probabilities is admissible as an interpretation of the probability calculus – let alone a procedure for measuring propensities. As a consequence, it has to be postulated that propensities behave like absolute probabilities. Now, one can imagine to have an analogous postulate in the conditional case. In other words, merely postulating that the proposed interpretation accounts for the properties of Bayesian conditionalization is a strategy available to a supporter of the propensity interpretation of absolute probabilities.

Acknowledgedly, my arguments against possible objections to the propensity interpretation of conditionalization that I have proposed are rather weak. However, I do not see the content of the proposed interpretation of conditionalization as the most interesting aspect of the present paper. Positively, what I consider as the core contribution of the paper to the debate concerning conditional propensities is precisely its contribution to redefining this debate. I have pointed out that Humphreys, as well as the other participants to the ‚debate consider that an interpretation of conditionalization is contained in Popper's proposition to interpret absolute probabilities as measures of propensities. On the other hand, I have shown that conditionalization requires its own *interpretation* – and I have given an analysis of what it formally is to give such an interpretation. Actually constructing the interpretation, then, is the last job. Maybe the way I have carried it out is not satisfactory; still, others may formulate more convincing propositions. It is my final contention, indeed, that there can be a propensity interpretation of conditional probabilities.

[63] [Lewis76] p. 311.

Annexe B

Extended abstract

Introduction

Introduction

It is clearly true that causality plays a central role in our explanations, be they scientific or not. It is also clearly true that the efficacy of our actions depends on our knowledge of causes. Yet it is false that we are done with causality. This does not mean that philosophers have failed to be interested in causality; on the contrary, causality has been a topic for philosophy all along its history. Rather this means that there exist lots of attempts at explicating causality, that it is not easy to understand how these various attempts hang together, and above all that none of them is generally accepted as correct.

Although it is quite difficult not to get lost in the area of attempts at explicating causality, one can identify a salient element of the recent development of this area. In order to understand that, one has to know that the idea of causation as a relation of necessitation has remained unchallenged (or very nearly so) until the middle of the twentieth century. In Anscombe's terms: "the truth of this conception is hardly debated. It is, indeed, a bit of *Weltanschauung*"[1]. More precisely, the beginning of Anscombe's paper explains how the idea that causation is a necessitation relation has gone through the whole history of occidental philosophy. In particular, Anscombe explains that Hume's criticism of the idea of causation as a *logical* relation did not untie causality and necessitation: "as touching the equation of causality with necessitation, Hume's thinking did nothing against but curiously reinforced it"[2].

Now, precisely this equation has been brought into question during the

[1] Anscombe (1981) p. 89.

[2] Anscombe (1981) pp. 89–90.

1960s. More clearly, *probabilistic* theories of causality appeared during the 1960s, and they developed out of two theses: first some causes do not make their effects necessary, but second il must be possible to characterize causes by their making their effects more probable. My work deals with those theories. More precisely, the questions I examine belong to the theoretical field which appeared when probabilistic theories of causality were first formulated.

Next subsection of the introduction sets out these theories, or more exactly the aspects of these theories that motivate my work. This leads to the idea that the distinction between two kinds of causes, generic and singular, is central as regards probabilistic theories of causality. As a consequence, the second subsection discusses the distinction between generic causes and singular causes. In the third and fourth subsections, I expose the problems which I will be interested in. Some are about generic causes, others about singular ones.

Probabilistic theories of causality

As already stated, introducing probabilistic concepts in the analysis of causality unties the ancestral link between causation and necessitation. More precisely, probabilistic theories of causality fall within a framework which were created by Hume's analysis of causality, and within this framework they untie causation and necessitation. For Hume, causality is before all regularity, and it is characterized in particular by constant conjunction: an effect invariably follows its cause. Yet Hume's regularity analysis is clearly not sufficient to characterize causality. Indeed it implies that striking a match does not cause its lightning if there exist situations in which striking a match is not followed by a its lightning. In more general terms, Hume's thesis cannot account for the fact that most causes are not sufficient for their effects, and that they produce them only when some other factors (also labeled "causes" by most theories of causality) are present. As concerns the match, one of those factors is the presence of oxygen in the environment where it is struck.

Solving this difficulty does not require that probabilities come into the analysis of causality. Indeed it is correctly treated in the framework of refined regularity analyses of causality, especially the one proposed in Mackie (1974). Mackie characterizes a cause as "an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition"[3] for its effect – a "condition" being here a set of factors.

Be they refined or not, regularity analyses of causality assume that there is no effect without a set of factors sufficing to produce it – and that the

[3]Mackie (1974) p. 62. Italics are in the original text.

effect follows regularly. On the other side, there is no cause but belonging to a set of factors sufficing to produce the effect. But it is not clear whether all causes are of this kind. In other words, it is not clear that there do not exist causes that produce their effect without belonging to a set of factors which suffices to this production. More precisely, one usually considers that the hypothesis of some causes not belonging to sets of factors sufficient for their effects took shape with the discovery of quantum phenomena and that it is very plausible today, including outside the quantum area. For example, we do not know of any set of factors to which the property of smoking would belong and which would be sufficient for one developing lung cancer. Most important, nothing guarantees that such a set exists. More generally, the hypothesis of effects without sets of factors sufficing to produce them is at least very plausible, if not established.

Contrary to regularity analyses, probabilistic theories of causality are compatible with this hypothesis. Indeed, they rely on the idea to characterize a cause as making its effect more probable. More explicitly, a cause C makes its effect E more probable in the sense that the conditional probability $p(E|C)$ has a value greater than that of the absolute probability $p(E)$. I'll explain later[4] that this idea has to be completed in order to serve as an analysis of causality. But for the time being it will be enough to underline the following: that a cause makes its effect more probable implies neither that it gives this effect probability 1, nor that it belongs to a set of factors that gives this effect probability 1. As a consequence, the idea on which probabilistic theories of causality rely is indeed compatible with the hypothesis of effects without sets of factors sufficing to produce them.

First among the questions raised by probabilistic theories of causality is the question of their status. On that point I stand up for three, non independent, theses. First, probabilistic theories of causality are *conceptual analyses*. They answer the question of what it means for A to cause B. In particular, it is only secondly, if ever, that they give criteria for effectively recognizing causes. Similarly, they are not definitions of causality. Indeed, my second thesis about probabilistic theories of causality is as follows: they are analyses of *one aspect* of the concept of causality. In other words, probabilistic theories answer *one* question concerning causality. This question could be given the following formulation: what is the co-occurrence relationship between causes and their effects? This differs in particular from the question of the kind of realities that causes, effects, and cause-effect relations are. Third and finally, probabilistic theories of causality first appear as theories of *generic* causation, rather than singular causation.

[4] In section B.2.

The problems I tackle largely depend on this last point. Indeed, generic causation and singular causation are in dissimilar positions with regard to probabilistic theories. This is shown in next subsection.

Generic causation and singular causation

The distinction. As far as propositions are concerned, the difference between generic causation and singular causation is exactly the difference between "Smoking causes lung cancer" and "Peter's smoking caused him to develop lung cancer", or between "Falls cause fractures" and "My falling in the stairs this morning caused my wrist to be fractured". Thus generic causation is a relation between properties – for example the properties of having a fall and of having a bone fractured. On the other side, singular causation is a relation between singular events which actually occurred. Generic causation and singular causation are usually referred to as *levels* of causality.

How levels of causality relate remains a debated question. The three leading answers are as follow: 1) singular causal relations are causal only because they instantiate generic causal relations, 2) generic causality gets its reality from singular causality, 3) singular causality and generic causality are independent. I will not side with any one of these answers. More generally, the question of the relationship between levels of causality is not tackled in my work.

The reason why I do not treat this question is that the answer it receives is neutral as regards what can be said of probabilistic theories of causality, in particular of their appropriateness and relevance. As a justification, consider the fact that each one of the three leading answers has been supported by some proponents of probabilistic theories of causality. The first answer is implied by Suppes' pretention to extend Hume's analysis to causes that do not suffice to produce their effects.[5] Indeed, for Hume, singular causal relations exist *as causal relations* only in as much as they instantiate generic causal relations.[6] The second answer is at work in Cartwright (1989) and in Humphreys (1989). The third answer is defended in Sober (1985) and in Eells (1991) with different consequences: Sober considers that only generic causation can be given a probabilistic theory, Eells develops two theories resorting to probabilities, the first one being a theory of generic causation and the second one a theory of singular causation.

I have just explained that there is no need for me to answer the question of the relationship between levels of causality. Still I cannot avoid the following

[5]Suppes (1970) p. 9.
[6]Hume (1739) p. 150.

remark: generic causation and singular causation are not in similar positions with regard to probabilistic theories. In other words, there are several reasons why there is no analogy between probabilistic theories of generic causation on the one hand and probabilistic theories of singular causation on the other hand. In the end of the present subsection, I explicate this thesis and I set out probabilistic theories of generic causation and probabilistic theories of singular causation to the extent that is needed for the problems I will tackle to appear.

Generic causation and probabilistic theories. In order to understand that generic causation and singular causation are not in similar positions with regard to probabilistic theories, one can come back to the thesis that probabilistic theories first appear as theories of generic causation. At that point the following precision must be given: here "first" is not to receive a temporal interpretation. Indeed, the analysis proposed by Suppes in the book which really founded the field of probabilistic theories of causality is meant to apply to singular causation as well as to generic causation.[7] Moreover, only around 1990 did probabilistic theories become explicitly presented either as theories of generic causation, or as theories of singular causation.[8]

What I meant when saying the probabilistic theories appear first as theories of generic causality is mainly that counter-examples to the idea that causes make their effects more probable concern the singular case. This is the case of Rosen's famous counter-example: a mediocre golf-player, Jones, shoots the ball, which hits a tree-limb, but hits it in such a way that it falls directly into the cup. Jones' shoot *causes* him to make a birdie although it diminishes the probability of a birdie. Most important, the causal relation between Jones' shoot and the birdie is specifically singular: it is *this* shoot by Jones' that causes the birdie, not mediocre Jones' shoots in general.

Moreover, I claim that the probabilistic theories of generic causation that are developed in Cartwright (1989)[9] and in Eells (1991)[10] are satisfactory. On the one hand, these theories give satisfactory accounts of all the cases which had been identified as problematic for preceding probabilistic theories of generic causation. On the other hand, and correlatively, the debate concerning probabilistic theories of generic causation seems to have dried up after those two books got published. According to the thesis I presently support, we now know what "smoking causes lung cancer" means.

[7] Suppes (1970) p. 75.
[8] Particularly: Cartwright (1989), Humphreys (1989), Eells (1991).
[9] Cartwright (1989) chap. 3 and 4.
[10] Eells (1991) part 1.

Singular causation and probabilistic theories. Things are appreciably different on the side of singular causation. I have just explained that the idea of characterizing causes as making their effects more probable is less adapted to singular causation than to generic causation. More precisely, we have seen that one particular shoot by Jones can cause him to make a birdie although it made this event less probable. Yet there exists at least one sound probabilistic theory of singular causation: the one that is proposed in Humphreys (1989).[11] Now this theory has three remarkable features that make it likely to be criticized, and imply that singular causation is not in the same situation as generic causation as regards probabilistic analysis.

First, Humphreys' theory sometimes leads to causal judgements that are counter-intuitive. Woodward in particular underlines this point.[12] It is true that Humphreys forestalls the objection by claiming that systematic philosophy should prevail over ordinary language in case they conflict.[13] Yet this argument has not been enough to convince that Humphreys (1989) makes tenable the idea of characterizing a singular cause by its making its effect more probable. Positively, the arguments developed both in Sober and Eells (1983) and in Sober (1985) to the effect that only generic causation is liable to probabilistic analysis largely keep being considered as correct.

Second, Humphreys' theory is a theory of causation between properties instantiations. In other words, the theory leads to consider that an event likely to be cause or effect always is the instantiation of a property by a system: "an event is a change in, or possession of, a property in a system on a trial"[14]. Thus, singular events are defined in reference to what they are not, and more precisely in reference to precisely the *relata* of generic causation. Against this, one can put forward that singular events are well and truly singular only when defined in reference to what they specifically are: really actualized in physical space and time. This requirement seems to be all the more legitimate as it is compatible with the now classic characterization of events by Quine: "Each [event] comprises simply the content, however heterogenous, of some portion of space-time, however disconnected and gerrymandered" .[15] Humphreys' theory at least lets unanswered the question of its application to singular events conceived along those lines.

Third, the probabilities through which Humphreys analyzes singular causation are physical singular probabilities.[16] This is a consequence of the na-

[11] Humphreys (1989) §31.
[12] Woodward (1994) pp. 366–367.
[13] Humphreys (1989) p. 5.
[14] Humphreys (1989) p. 24.
[15] Quine (1960) p. 171.
[16] Humphreys (1989) p. 54 in particular.

ture of the object to be analyzed: causation in as much as it is a feature of the world.[17] On the other hand, this results in the probabilities used by Humphreys to analyze singular causation being given a propensity interpretation. Indeed, among interpretations of probability, the propensity one is the only one that makes sense of the notion of physical probabilities of singular events. Yet if the propensity interpretation is the only one to make sense of physical probabilities of singular events, it is also the only one for which we have serious doubts that it makes sense of *conditional* probabilities. However, it appeared earlier on that the idea of probability raising on which probabilistic theories of causality rely, is to be understood in terms of conditional probabilities. There appears a third aspect of Humphreys' theory that may be criticized.

As already explained, the questions I treat in this work are raised by probabilistic theories of causality as they are developed today. But I have shown that generic causation and singular causation are not in similar positions as regards those theories. Correlatively, probabilistic theories raise very different questions depending on whether one is interested in generic causation or singular causation. As a consequence, my work divides into two parts. The first one corresponds to the first four sections of the present appendix and deals with questions that are opened by probabilistic theories of generic causation. The second part, corresponding to sections B.5 and B.6 of the present appendix, deals with questions related to probabilistic theories of singular causation.

Probabilistic theories and the epistemology of generic causation

If granted that probabilistic theories are satisfactory analyses of the concept of generic causation, one naturally comes to the problem of criteria that make generic causes recognizable. This, at least, is suggested by the distinction between analyzing causality, giving a criterion of effective recognition of causality, and defining causality. This distinction was introduced earlier on and I then supported that probabilistic theories are essentially conceptual analyses and that they deal with only one aspect of causality. As a consequence, there can be no probabilistic definition of generic causality. Then, the only question that remains unanswered if probabilistic theories of causality are correct is the epistemological one: how are we to recognize generic cause-effect relations?

[17] Humphreys (1989) pp. 54–55.

This question is not justified only in a negative way and by the existence of an untreated problem. Positively, the question is justified by the relationship between conceptual analysis and criteria for effective recognition: having an analysis of "A generically causes B" is a good starting-point toward the formulation of criteria for recognizing generic causes. The question, then, becomes more precise: do probabilistic theories of causality give usable criteria for recognizing generic causes? Put methodologically: are there methods for inferring generic causes that resort to recognition criteria inherited from probabilistic theories of causality?

Here, causal inference methods relying on Bayesian networks appear as both good candidates and a matter deserving of our attention. These methods, coming from artificial intelligence, appeared at the turn of the 1990s. Essentially, they consist in algorithms constructing a graph when given probabilistic information of the same kind as statistical information. This graph is directed and the arrows in it are causally interpreted: each arrow is taken to represent a cause-effect relation. This first, very abstract, description makes it clear that the causal inference methods I refer to use a probabilistic criterion for causality. Now this criterion could be compared with probabilistic theories of causality. The idea of such a confrontation looks all the more relevant since the criterion used by Bayesian networks causal inference methods is clearly related to probabilistic theories of causality. As a first justification, consider the fact that it is not rare that surveys of theories of causality make this criterion a particular probabilistic theory. Hitchcock (2002) is an example.

The question, then, arises of the exact relationship between on the one hand the criterion for causality that is used by causal inference methods resorting to Bayesian networks and, on the other hand, probabilistic theories of generic causation. This question arises all the more naturally that the criterion for causality that Bayesian networks convey is easily identified. As conceded by Cartwright, Bayesian networks causal inference methods "are the most explicitly and carefully grounded methods for causal inference available"[18]. Indeed Bayesian networks are defined precisely by the relationship between probabilities and causally interpreted graphs. What this relationship exactly is and whether it actually holds is examined in section B.1[19]. In more general terms, this section is devoted to present Bayesian networks and discuss their causal interpretation.

[18]Cartwright (1999) p. 20.

[19]More precisely, this is examined in chapter 1 of the French thesis and section B.1 is the abstract for this chapter. More generally, there is a one to one correspondence between the sections of the present extended abstract and the chapters of the French thesis. How those sections relate to these chapters will not be mentioned anymore.

Now let us come back to the more specific question of the relationship between probabilistic theories of causality and causal inference as it is allowed by Bayesian networks. This question is all the more urgent since probabilistic theories of causality do not immediately give rise to effective criteria for inferring causes. Indeed, in their completed form, probabilistic theories are rather complex and they are circular, in the sense that the analysis of the concept of cause resorts to causal concepts. As a consequence, one wonders why and how Bayesian networks allow causal inference. Is it that they use a criterion for causality that is not, in the end, parent of probabilistic theories? Or that complexity and circularity do not prevent probabilistic theories of causality from giving a recognition criterion that would be effectively usable? In this last case, what role does the graphic component of Bayesian networks play regarding the possibility to actually infer causes? These questions are tackled in section B.2.

Due precisely to this graphic component, causal Bayesian networks fall within a tradition that was opened by Wright in the 1920s. This tradition is that using causal models for the epistemology of generic causality. The proponents of Bayesian networks claim that their specificity as causal models is that they enable *induction* of causes. If this is true, then Bayesian networks considerably renew the epistemology of generic causation, which is traditionally attached to hypothetico-deduction. Is this really so? Are causal inference methods based on Bayesian networks really inductive? More generally, how do Bayesian networks contribute to generic causal inference methodology, and how do the methods based on Bayesian networks differ from the more traditional one with which they compete? These questions are treated in section B.3.

The questions that are treated in section B.3 appear more important still when considered as sub-questions of the following general question: can causal knowledge be inferred from probabilistic data, and how? This question has a story which dates back to the end of the nineteenth century at least, when French sociology was constituted. Indeed, methodological considerations proved essential to sociology becoming free from psychology, and one question which received particular consideration is precisely whether it is possible to use statistical data in order to get causal conclusions.[20] Thus, section 3 comes back to classical questions, but offers them a treatment in the new light that is shed by Bayesian networks.

In a noticeably different way, section B.4 of the present appendix comes straight to some of the more contemporaneous questions that are raised by causal Bayesian networks. Basically, these are questions about the satisfac-

[20]Durkheim (1895) is an illustration of both points. See the last chapter in particular.

tion of the hypotheses that ensure that Bayesian networks causal inference methods can be resorted to. More precisely still, the point is to characterize the real systems that satisfy these hypotheses. As far as the "causal Markov condition", which is probably the most important of these hypotheses, is concerned, it first looked as if it were more likely to be satisfied by deterministic systems than by indeterministic ones. This thesis was supported by precisely those who introduced and now defend Bayesian networks causal inference methods. It had never looked problematic until Steel (2005), which puts it into question. Now is determinism more favorable than indeterminism for the causal Markov condition? Section B.4 is devoted to answer this question. Then I come to the second part of my work, which I present now.

Probabilistic theories of singular causation

I have already explained that generic causation and singular causation are not in similar situations regarding probabilistic analysis. More precisely, I have pinpointed three aspects of our best probabilistic theory of singular causation – the one developed in Humphreys (1989) – that make it debatable. As a consequence, in the singular case, probabilistic theories of causality as they stand today raise questions that are different from the ones that have just appeared in the generic case. More explicitly, the second part of my work does not tackle questions that would be analogous to the epistemological questions that will interest me in the first part of my work. Positively, the second part sticks to conceptual analysis, that is to the discussion of what it means for A to be a singular cause of B, of what it means that Jones' shoot caused the birdie, or that Peter's heavy smoking caused him to develop lung cancer.

This question is broached starting from two of the three remarks that I have formulated concerning Humphreys (1989). The first of these two remarks is the one about events, and more precisely about whether it is accurate to define events by reference to properties. I have already claimed that characterizing an event as the content of a spatiotemporal zone better grants what makes singular events specific, and hence what contributes to distinguish singular causation from generic causation. The then coming out project is the project of a probabilistic analysis of what may be called "actual causation", that is singular causation considered as a relation between events defined in terms of what really makes them singular. Before such an analysis may be produced, it has to be explained what the relationship between actual causation and probabilities is. The second part of my work deals with this question, and tries to supply with some elements of an answer.

The question of the relationship between actual causation and probability

is approached from the philosophy of probability. This approach is justified by its enabling to determine the still vague question of the relationship between actual causation and probabilities. More precisely, it is my contention that probabilities bear a relationship with actual causation – and, beyond that, can serve to analyze it – only if they are interpreted in terms of propensities. Indeed, the propensity interpretation is the only one of the available interpretations of probability that makes sense of probabilities of singular events that are features of the physical world in the same way as actual causation is. As a consequence, it is true that one determines the question of the relationship between actual causation and probabilities when approaching it from the philosophy of probability. More explicitly, the question becomes that of the relationship between on the one hand causation, and on the other hand probabilities given a propensity interpretation.

This last question is meaningful by itself, that is independently from the question of the relationship between actual causation and probabilities. Indeed, the propensity interpretation of probability as introduced by Popper in the 1950s relies on the idea that probabilities measure physical dispositions to produce singular events, these dispositions being called "propensities". Now, under the realist conception towards which Popper often tends, propensities are surprisingly like causes: they are physical entities capable of making events happen. Are propensities causes, and in what sense? What does ensue for epistemology, ontology, and even metaphysics? These questions are tackled in section B.5. This section was conceived as an in-depth presentation of the propensity theory of probability in the Popperian tradition. This means that the philosophical correlates of the theory are taken into account, and also that the questions tackled are of more general philosophy than the questions usually attached to the sole philosophy of probability.

In section B.5, I confine myself to the propensity theory of probability as an interpretation of *absolute* probabilities. But I have already explained that explicating the idea on which probabilistic theories of causality are grounded – that a cause makes its effect more probable – makes *conditional* probabilities appear. It follows that the second part of my work can be what I have pretended it would be only if I do not stick to the absolute case, and turns on the relationship between causation and conditional probabilities interpreted in terms of propensities.

That is where the third of the remarks I formulated earlier on concerning Humphreys' probabilistic theory of singular causation enters. More precisely the remark was that Humphreys resorts both to conditional probabilities and to propensities when there exist serious objections to the idea of a propensity interpretation of conditional probabilities. I have considered this as the basis for a possible criticism of Humphreys' theory. Now it can be considered a

threat to my project of extending the analyses in section B.5 to conditional probabilities. The threat will appear all the more serious since section B.5 will insist on the causal dimension of the Popperian propensity interpretation of absolute probabilities, and that this causal dimension leads to the difficulties that the propensity theory has with conditional probabilities. To put it in comparative terms: the propensity theory that is presented in section B.5 is much more exposed than the one supported in Humphreys (1989), to the difficulties raised by conditional probabilities.

The consequence seems to be that I have to leave the project of clarifying the relationship between actual causation and conditional probabilities. More exactly, the project has to be left unless it is possible to get round the obstacle that conditional probabilities are for propensity theories of probability. Section B.6 is very largely devoted to explore the ways of possibly getting round this obstacle. In this section, I discuss the reasons why we think that the propensity theory cannot be an interpretation of conditional probabilities, and I propose what could be a propensity interpretation of conditional probabilities. What ensue from this proposition regarding the relationship between actual causation and conditional probabilities? How does this link up with the analyses in section B.5, which concern the relationship between causation and absolute probabilities given a propensity interpretation? These questions are treated towards the end of B.6.

B.1 Causal Bayesian networks

This section is a presentation of causal Bayesian networks. As our interest in sections B.1 to B.2 is in causal inference as based on Bayesian networks ("Bayes nets" from now onwards), the focus is on the limitation that the very notion of causal Bayes nets imposes on Bayes nets causal inference. The section divides into three subsections. First, I set out Bayesian networks as such, that is independently from their causal interpretation. Second, I come to this interpretation and I explain why it seems plausible, to say the least. Third, I qualify this analysis and discuss usual counter-examples to the causal interpretation of Bayes networks.

B.1.1 Bayesian networks

B.1.1.1 Where do Bayesian networks come from?

Understanding what Bayesian networks are requires to have an idea of where they come from. As it is clear that they emerged in artificial intelligence at the beginning of the 1980s and in relation to the representation of uncertainty, I first set Bayes nets in the field of artificial intelligence representations of uncertainty. This can be made through a comparison of the treatment of uncertainty that is conveyed by Bayes nets, with the treatment that is at work in an expert system such as MYCIN.

Both treatments can be characterized as *numerical* and, as such, opposed to the *logical* treatment corresponding to non-monotonic logic. But this convergence must not hide two important differences:

- the numerical concept used by MYCIN is not probabilistic[21], whereas classical probabilities are essential to Bayes nets;
- this numerical concept is handled through local syntactic rules in MYCIN whereas Bayes nets imply global, semantic updating.

Treating uncertainty semantically has computational drawbacks. These drawbacks can be overcome only if one finds a way to efficiently represent what can be ignored when updating a given probability. This is exactly what the graphic component of Bayes nets does.

B.1.1.2 Definitions

Once explained where Bayes nets come from, I can come to their definition. As a Bayes net is composed of a graph and a probability distribution, defining

[21] Shortliffe et Buchanan (1984) p. 239.

Bayes nets requires notions from graph theory and from probability theory. The fundamental definitions are given in the appendix to chapter 1. They enable to define Markovian parents:

Definition B.1.1 (Markovian parents) *Let $\boldsymbol{V}$ be a set of variables*[22]*, V_i in $\boldsymbol{V}$, $<$ a strict order over $\boldsymbol{V}$ and p a probability distribution over $\boldsymbol{V}$.*
A set $\boldsymbol{PM}_i$ of Markovian parents *for V_i in $\boldsymbol{V}$ with regard to p and $<$ is a subset of $\boldsymbol{V}$ that is minimal among those that have the following properties:*

- *for each V_j in $\boldsymbol{PM}_i$, $V_j < V_i$;*
- *conditional on $\boldsymbol{PM}_i$, V_i is independent from all its other $<$-predecessors in $\boldsymbol{V}$.*

The idea leading from Markovian parents to Bayes nets is as follows: representing sets of Markovian parents as parents in a directed acyclic graph. In order to make this idea rigorous, two more definitions are needed:

Definition B.1.2 (Agreement between $<$ and G) *Let $\boldsymbol{V}$ and $<$ and before, and G a directed acyclic graph over $\boldsymbol{G}$.*
$<$ agrees with G *if : for all V_i and V_j in $\boldsymbol{V}$, if V_i is an ancestor of V_j in G, then $V_i < V_j$.*

and:

Definition B.1.3 (Representation of p by G) *Let $\boldsymbol{V}$, p and G be as before, and for each V_i in $\boldsymbol{V}$ let $\boldsymbol{PA}_i$ be the set of its parents in G.*
G represents p *if : for all V_i in $\boldsymbol{V}$, $\boldsymbol{PA}_i$ is a set of Markovian parents of V_i with regard to p and any strict order $<$ over $\boldsymbol{V}$ that agrees with G.*

Now Bayesian networks can be defined as follows:

Definition B.1.4 (Bayesian network) *Let $\boldsymbol{V}$ be a set of variables.*
A Bayesian network *over $\boldsymbol{V}$ is (G, p) such that:*

1. *G is a directed acyclic graph over $\boldsymbol{V}$;*
2. *p is a probability distribution over $\boldsymbol{V}$;*
3. *G represents p.*

[22] All the sets of variables that are considered are finite sets of variables each of which has a finite number of values.

B.1.1.3 Two results concerning Bayesian networks

There exist two important results concerning Bayes nets thus defined. The first one gives a condition equivalent to G representing p:

Definition B.1.5 *Let $\boldsymbol{V}$ be a set of variables, G a directed acyclic graph over $\boldsymbol{V}$ and p a probability distribution over $\boldsymbol{V}$.*
G represents p if and only if each variable in $\boldsymbol{V}$ is independent for p of all its non-descendants in G conditional on its parents in G.

The condition to which representation of p by G is equivalent is known as the "causal Markov condition" and it can be used to define Bayesian networks.

The second result concerning Bayes nets gives a graphic criterion – called "d-separation" – for conditional probabilistic independence. More precisely, if G is a directed acyclic graph over $\mathbf{V}$, then d-separation in G is equivalent to conditional probabilistic independence for all the probability distributions that are represented by G.

B.1.1.4 Uses of Bayesian networks

Due to the two results that have just been mentioned, Bayes nets as they are defined in the present subsection – that is uninterpreted Bayes nets – can serve at least two purposes. First, Bayesian networks can serve as an economical definition of probability distributions. Indeed, it comes straightforwardly from the definitions given in B.1.1.2 that a probability distribution p over $\mathbf{V} = \{V_1, V_2, \ldots V_n$ is completely defined by:

1. a directed acyclic graph that represents p – that is a graph G such that (G, p) is a Bayesian net;

2. the conditional probabilities $p(v_i|\ \mathbf{pa}_i)$, where v_i is a value of V_i belonging to $\mathbf{V}$ and $\mathbf{pa}_i$ is a value of the set $\mathbf{PA}_i$ of V_i's parents in G.

This definition of p is economical in the sense that the numbers of parameters that have to be specified in 2. is smaller than the number of parameters that would have to be specified in the absence of G. More precisely, it is strictly smaller as soon as there exist two variables in $\mathbf{V}$ that are not adjacent in G – and the sparser G, the more important the difference between the definition of p with G and the definition of p without G.

Second, uninterpreted Bayesian networks can be used to update probabilities. More exactly, a probability distribution that is defined along the just exposed lines is more easily updated than a probability distribution defined otherwise. Indeed, G indicates which variables must be taken into account

when calculating the up-dated values of a given variable. Conversely, G indicates what can be ignored when calculating these values, and this makes calculation tractable. Beginning with Pearl (1982), more and more general algorithms for up-dating probabilities in Bayes nets have been proposed.

B.1.1.5 Bayesian graphs

Uses of uninterpreted Bayes nets heavily rely on their graphic component – that is on Bayesian graphs. Concerning precisely these graphs, first it is easy to show the following: the information carried by a Bayesian graph can be expressed in natural language and there exists a mechanical procedure for finding a natural language expression of the information conveyed by a given Bayesian graph.

In spite of this Bayesian graphs are very useful. There are two reasons why they are: given a set of variables $\mathbf{V}$, a probability distribution p over $\mathbf{V}$, a probability distribution and a strict order over $\mathbf{V}$ and sets $\mathbf{PA}_i$ of Markovian parents of the variables in $\mathbf{V}$ relative to p and $<$:

1. the conditional independencies for p that derive from the $\mathbf{PA}_i$s can be *read* on a graph G representing p;

2. the conditional independencies for p that *do not* derive from the $\mathbf{PA}_i$s can also be read on G.

B.1.2 Causally interpreted Bayesian networks

Once the uninterpreted notion of Bayes nets has been presented, one can come to the causal interpretation. A causal Bayesian network is a Bayesian network whose graph is interpreted causally, which means that an arrow between two variables of the graph is taken to represent a direct cause-effect relation. The property, for a cause-effect relation, to be direct is relative to the set of variables that is considered.

The causal interpretation of Bayes nets relies on two hypotheses. The first one is about causality itself and is follows: direct cause-effect relations among the variables of a given set can be represented by a directed acyclic graph over this set. This I call the *representation hypothesis*. Second, the causal interpretation of Bayes nets supposes that the direct causal graph over a set of variables and the probability distribution over that set are related as held by the Markov condition. This is the *causal Markov condition*. In the present subsection, I analyze these hypotheses and show that they are plausible. Before that, however, I set out the contexts in which the notion of causal Bayes nets appear.

B.1.2.1 Uses of causal Bayesian networks

The notion of causal Bayes net appears in two contexts. In both cases, the fact that the representation hypothesis and the causal Markov condition are plausible is essential. First, this implies that the graph representing the direct cause-effect relations among the variables of a given set represents the probability distributions over this set. Here, direct causation is known and as a guide for constructing Bayes nets. This is how causality was originally related to Bayes nets.

Second, the plausibility of the representation hypothesis and causal Markov condition leads to the idea that the algorithms that build the graph representing a given probability distribution, in fact build a causal graph. Here causality is *inferred* – hence this is the context that interests me in the first part of my work. Notice that the algorithms I mention rely not only on the causal Markov condition, but also on the Faithfulness hypothesis:

Definition B.1.6 *Let G be a directed acyclic graph over a set of variables $\boldsymbol{V}$, and p a probability distribution over $\boldsymbol{V}$.*
(G, p) satisfies Faithfulness if there are no probabilistic independences for p other than the ones stemming from (G, p) being a Bayesian network.

Notice also that the mentioned algorithms do not output a single directed acyclic graph, but a pattern representing all the directed acyclic graphs that represent the given probability distribution.

B.1.2.2 The representation hypothesis

The representation hypothesis can be decomposed as follows:

1. causal *relata* can be represented by variables;
2. direct causation among the variables in a given set can be represented by a binary relation;
3. the graph representing direct causation among the variables in a given set is acyclic.

As far as 1. is concerned, it is plausible in the generic case. Indeed, every property can be represented by the variable taking value 1 when the property is instantiated and value 0 otherwise. Yet this is only a particular case. In general, a variable can represent a family of properties such that exactly one of them is instantiated by any individual belonging the population under consideration. This first sub-hypothesis deals with the granularity of the representation of causality, rather than it is substantial. The mode of

representation it conveys is coarse: an arrow between two variables does not tell which value(s) of the first one cause(s) which value(s) of the second one.

As far as 2. is concerned, it looks debatable in the context of a probabilistic approach of causality. Indeed probabilistic relations vary from populations to populations, and it has been claimed[23] that this leads to generic causation being a ternary relation: cause-effect relations are relative to a population. However, the fact that the arrows of a Bayesian graph are between two variables only does not exactly imply that causal Bayes nets impose binary causation. Indeed a Bayes net is composed of a graph *and a probability distribution*. If the probability distribution is relative to a sub-population, then so is causality as it is represented by the causal Bayes net. As a consequence, sub-hypothesis 2. is compatible with causality being considered a ternary relation; it can be regarded as non substantial.

As far as 3. is concerned, under the hypothesis that causation is the transitive closure of direct causation, it is equivalent to the asymmetry of causality. As a consequence, it inherits the plausibility of the thesis according to which causation is asymmetric.

B.1.2.3 The causal Markov condition

The causal Markov condition (CMM) states that, in a set of variables, a given variable is probabilistically independent of all its non-effects conditional on its direct causes. In other words, the causal Markov condition implies that, conditional on its direct causes, a variable is independent from any variable:

0. that is one of its direct causes;
1. with which it has no causal relationship;
2. that is one of its non-direct causes;
3. that is one effect of one of its direct causes, but not one of its effect.

Clearly, independences of type 0. are trivial.

Independences of type 1. can be taken to correspond to the scientific requirement that variations be explainable: no probabilistic dependence without a causal relation explaining it.

Independences of type 2. are acquainted to the idea of contiguity of causality, that is to the idea that causes are contiguous to their effects. I consider this idea true. But it is an idea concerning singular causation and not involving probabilities. However, it can be considered a particular version

[23]For example by Eells in Eells (1991) part 1.

of the idea that causation propagates step by step. To the extent that this general idea inherits the plausibility of the idea of singular causes being contiguous to their effects, independences of type 2. appear plausible.

Independences of type 3. have to do with Reichenbach's Principle of the common cause (PCC). More precisely, there exist two differences between independences of type 3. and the independences implied by the PCC. First, independences of type 3. are relative to *direct* causes. However, in a causal Bayes net, if there exists a cause common to V_1 and V_2 that screens off V_1 from V_2, then there exists a direct cause of V_1 that does it. In other words, in a causal Bayesian net, the independences corresponding to the PCC imply independencies of type 3. These independences inherit the plausibility of the PCC. Yet not all independencies of type 3. are implied by the PCC. Indeed, the second difference between independences of type 3. and independences implied by the PCC is that the first ones are relative to *sets* of variables, rather than to variables. This, however, makes the independences more liable to hold. As a consequence, independences of type 3. appear still more plausible.

To put it in a nutshell, the very notion of causal Bayes net conveys:

- one hypothesis concerning the granularity of the representation of causality: causal *relata* can be represented by variables;
- two hypotheses concerning causality: first it is asymmetric, second it is related to probability as stated by the CMC. I have explained why these two substantial hypotheses are plausible. It is time now to come to the counter-examples.

B.1.3 Counter-examples

B.1.3.1 Counter-examples to the asymmetry of causation

Our best argument in favor of the asymmetry of causation is temporal. More exactly it runs as follows: (whatever the status of this) causes precede their effect in time, hence effects cannot cause their causes. The matter is that this temporal argument holds for *singular* causation. On the contrary, it is not clear what it could mean in the generic case. Moreover, generic causation is not asymmetric: poverty causes crime which causes poverty, a weak immune system causes illness which causes a weak causes system...

However this does not prevent from representing the direct cause-effect relations among the variables of a set $\mathbf{V}$ by a directed acyclic graph G over $\mathbf{V}$. More precisely, it prevents this representation only if one considers that each variable belonging to $\mathbf{V}$ must appear exactly once in G. If this is not

the case, as explained by Williamson[24] causal cycles can be spread out. Then acyclicity is artificially restored.

B.1.3.2 Counter-examples to the causal Markov condition

Counter-examples to the causal Markov condition are mainly counter-examples to the independences of type 3. above, that is examples of causes that fail to make their effects independent. Such causes have been characterized by Salmon through the notion of interactive fork. They can also be used to build counter-examples to independences of type 1. As far as independences of type 2. are concerned, the only counter-examples I can envisage are built out of processes that are usually called "non-Markovian". However, a closer look at them reveals that they give counter-examples to the *temporal* Markov condition rather than to the *causal* Markov condition.

Fortunately, one can define a trick analogous to the one that restores acyclicity. This requires to accept that the graph representing the direct cause-effect relations among the variables of $\mathbf{V}$ is not necessarily a graph over $\mathbf{V}$, but may include variables not belonging to $\mathbf{V}$ and that may be unobservable. With such "hidden nodes", one can always restore the causal Markov condition.

B.1.3.3 Can the counter-examples be actually overcome?

Finally, I turn to whether the tricks that have been set out can be effectively used to make causal graphs acyclic and to restore the causal Markov condition true. Obviously, this depends on the context where causal Bayes nets appear. If causality is used as a guide for constructing Bayesian graphs, then surely the tricks I have presented can be resorted to. How they can be is more precisely stated in Gillies (2002).

However, the tricks cannot be resorted to if causality is what one pretends to infer when building a Bayesian graph. Indeed the outputs of the algorithms that build the graphs representing a given probability distribution over $\mathbf{V}$ are directed acyclic graphs *over* $\boldsymbol{V}$ and such that each variable in $\mathbf{V}$ appears *exactly once*. In case acyclicity and the causal Markov condition are not satisfied by the graph that represents the direct cause-effect relations over $\mathbf{V}$, then the output of the algorithms does not adequately represent these relations.

As already explained, the uses in which I am interested in the first part of my work are precisely of the second kind. I have just made clear what the status of the acyclicity hypothesis and of the CMC is in this context. I can

[24] Williamson (2005) p. 50.

now turn to the way causes are inferred using Bayes nets. The first question, then, is how the criterion for generic causes that causal Bayes nets convey relates to probabilistic theories of generic causation. Next section tackles this question.

B.2 Causal Bayesian networks and probabilistic theories of causality

In this section, I come to the question of the relationship between the criterion for generic causation that is conveyed by causal Bayes nets and probabilistic theories of causality. The question is justified as part of the project of exploring the epistemological correlates of probabilistic theories of causality. Yet it has also a specific justification: as already mentioned, our best probabilistic theories of causality are circular, and this makes it surprising that causes can be inferred using probabilistic data.

In subsection B.2.1, I bring into light the probabilistic criterion that is used to infer generic causes from probabilistic data using Bayes nets. From now onwards, this criterion will be called "BN criterion for causality". In subsection B.2.2 I discuss probabilistic theories of causality more closely than I did in the introduction. In subsection B.2.3 the comparison between the BN criterion for causality and probabilistic theories of causality is actually carried out.

B.2.1 BN criterion for causality

In order to bring into light the BN criterion for causality, one can either come back to the hypotheses that define causal Bayes nets, or directly examine Bayes nets causal inference algorithms. More precisely, one can look either for the probabilistic correlates of an arrow in a Bayesian graph over a set of variables satisfying the hypotheses required for BN causal inference, or for the probabilistic conditions for the algorithms to trace an arrow between two variables. I take up the first method for two reasons. First, it avoids focusing on *sufficient* conditions for *direct* causation, and thus makes easier the comparison with probabilistic theories of causality. Second, it enables to keep distinct two quite different epistemological matters: on the one hand the analysis of causality that underlies causal inference, on the other hand the methods for inferring causes.

As already explained, causal inference based on Bayes nets requires the satisfaction of three hypotheses: Acyclicity, Faithfulness and the CMC. When looking for a probabilistic criterion for causality, one has to focus on the latter two. Taken together, they convey the following characterization for causality:

Proposition B.1 (BN characterization of causality) *X causes Y in $\boldsymbol{V}$ if and only if X and Y are probabilistically dependent relative to the set of X's direct causes in $\boldsymbol{V}$.*

Here, X and Y are two variables in **V**.

BN characterization of causality has one consequence that will reveal important: if X causes Y in the BN sense, then X and Y are probabilistically dependent (meaning dependent relative to the empty set). This is easily shown resorting to *d*-separation – and more precisely to the second of the results that are presented in paragraph B.1.1.3.

B.2.2 Probabilistic theories of causality

The present subsection is devoted to set out probabilistic theories of causality. The presentation is not exhaustive: I explain only what is necessary for the comparison with BN characterization. The strategy adopted for the presentation is roughly historical. More precisely, I begin with the seminal idea for probabilistic theories of causality – that a cause may be characterized by its making its effects more probable – and I explain how successive probabilistic theories of causality take into account wider and wider classes of counter-examples to this idea.

B.2.2.1 The seminal idea

It can stated as follows:

Proposition B.2 (Seminal idea) *A causes B if and only if* $p(B|A) > p(B)$.

Here, A and B are properties. Moreover one easily show that $p(B|A) > p(B)$ is equivalent to $p(B|A) > p(B|not-A)$ (provided, of course, $p(not-A) \neq 0$.[25]

The seminal idea has never been considered sufficient to characterize causality. More precisely, two cases in which probability raising does not correspond to causation – that is two kinds of spurious correlations – are taken into account even by the first probabilistic theories of causality.

B.2.2.2 Two kinds of spurious correlations

In the present paragraph, I set out the two kinds of spurious correlations that are taken into account by all probabilistic theories of causality. The first one consists of correlations between effects and causes. Here the point is that probability-raising is symmetric. In other words, A raises the probability of B if and only if B raises the probability of A. As a consequence, if causes well and truly make their effects more probable, then effects make their

[25]The fact that conditioning properties must have probabilities different from 0 is not tackled as such, and in the sequel I omit the precision.

causes more probable too. However, even though generic causation is not asymmetric (see paragraph B.1.3.1), it is false that all generic effects cause their causes.

Second, effects of a common cause raise each other's probability. As an example, one can take two effects of smoking: developing cardiac diseases on the one hand and having yellowed fingers on the other hand. Each of them makes the other more probable. However, none causes the other. More generally, it is false that effects of a common cause are always in cause-effect relation.

These two kinds of spurious correlations are taken into account by Suppes when formulating the theory that opens the area of probabilistic theories of causality. Indeed, the theory is as follows:

Proposition B.3 (Suppes (1970)) *A causes B if and only if:*

1. *A is prior to B;*
2. $p(B|A) > p(B)$*;*
3. *there does not exist any C prior to A such that* $p(B|AandC) = p(B|C)$.

Clause 1. aims at identifying as spurious the correlations between effects and causes. Clause 3. aims at eliminating as spurious the correlations between effects of a common cause.

B.2.2.3 Problems with Suppes' theory

I identify five classes of counter-examples to Suppes' theory. These five classes can be reduced to three problems. First, Suppes' theory of causality does not provide a satisfactory treatment of the spurious correlations it takes into account. On the one hand, clause 3. is not enough to identify as spurious all spurious correlations that are due to a common cause. Indeed, and as already explained, there exist forks that are not conjunctive, but rather interactive. In such cases, common causes do not render their effect independent. This point being made, the analyses to come deal only with the non-interactive case. On the other hand, clause 1. does not give a satisfactory treatment of spurious correlations between effects and causes. Indeed it is not clear what it means for a property to be prior to another property. Above all, even if this could be made clear, clause 1. implies that generic causation is asymmetric – which was shown to be false in paragraph B.1.3.1.

The second problem with Suppes' theory is that it does not take into account all kinds of spurious correlations. On the one hand, it does not take into account the possibility of spurious correlations that 1) are implied by *several*

common causes and 2) do no disappear by conditionalizing on any *one* of these causes. On the other hand, it does not take into account the possibility of spurious correlations corresponding to instances of Simpson's paradox. In such cases, probability-raising in one population hides probability-lowering in possibly all its sub-populations. Notice that the correlation in the population is not spurious in the sense as the spurious correlations that have been envisaged hitherto. Indeed, the problem is not that some probabilistic dependencies have no causal counterpart, but that some probabilistic relations apparently lead to identify the wrong properties as causally related.

The third problem with Suppes' theory is that it does not take into account the possibility of spurious independencies, that is the possibility that probabilistically independent properties are related as cause and effect. One kind of spurious independencies is provided by instances of the Simpson's paradox. There exist more complicated cases, but it is not necessary for me to take them into account in the present section. Clause 2. is responsible for Suppes' theory not giving spurious independencies a proper treatment.

B.2.2.4 Probabilistic theories posterior to Suppes (1970)

Probabilistic theories of causality that follow Suppes' one take into account both the spurious correlations and the spurious independencies that are implied by Simpson's paradox. This is done by restricting the requirement of probability-raising to some sub-populations of the population under consideration. Examination of instances of Simpson's paradox makes it clear that the relevant sub-populations must be characterized in a causal.

Depending on whether causality is taken to be probability-raising is all sub-populations or in one sub-population, one gets either the theory proposed in Cartwright (1979):

Proposition B.4 (Cartwright (1979)) *A causes B is and only if $p(B|A.\boldsymbol{S}_i) > p(B|\boldsymbol{S}_i$ for any $\boldsymbol{S}_i$ that is a state description over the set of the properties that cause B and are not caused by A.*

or the theory proposed in Skyrms (1980):

Proposition B.5 (Skyrms (1980)) *A causes B is and only if:*

1. *there exists j such that $p(B|A.\boldsymbol{S}_j) > p(B|\boldsymbol{S}_j$;*
2. *there does not exist k such that $p(B|A.\boldsymbol{S}_k) < p(B|\boldsymbol{S}_k$*

where the $\boldsymbol{S}_i$ are the state descriptions over the set of the causes of B that are not caused by A.

In both cases, the account is circular in the sense that the notion of cause appears in the *analysans* for "A causes B".

As for the relevance of these two theories of causality, I have already explained that spurious correlations and spurious independencies that are stem from Simpson's paradox are correctly treated by construction. It is easily shown that so are spurious correlations between effects and causes, and spurious correlations between effects of *several* causes. However, interactive forks remain problematic, as well as spurious independencies that are not related to Simpson's paradox. Posterior theories of causality offer correct treatment of these cases. Given my project of the present section, it will be sufficient that I indicate that those theories are also circular in the sense that has just been defined. In other words, what I have already stated will be enough for me to compare BN characterization with probabilistic theories of causality.

B.2.3 Comparison

The comparison at which the present section aims can now be led. However, one immediately faces the fact that BN characterization and probabilistic theories of causality do not have exactly the same objects. As a consequence, the strategy I adopt is as follows. First, I compare the objects of both analyses and above all show how the differences between these objects can be reduced. This leads me to produce an object that is liable to BN characterization as well as characterization by probabilistic theories. Second, precisely on this object, I compare the analyses properly speaking. Third, I show how the comparison sheds light on the very possibility to infer generic causes with Bayes nets.

B.2.3.1 Comparison of the objects

The objects of BN characterization and probabilistic theories differ in two remarkable ways. First, BN characterization deals with causality *relative to a set of variables* (see proposition B.1), whereas probabilistic theories are about causality (full point). This difference is noticed in Spohn (2001), which also indicates the way of its reduction. More explicitly, Spohn proposes to consider causality full stop as causality relative to a set of variables that suffices to describe empirical reality. I adopt this solution. In particular, I claim that the difficulties apparently raised by the notion of a set of variables that suffices to describe empirical reality are not problematic in the context of conceptually analyzing causality.

Second, BN characterization is about causality *between variables* whereas probabilistic theories deal with causality *between properties*. I have already explained in section B.2 that this feature of BN representation of causality makes it coarse. Moreover (and most important), I have indicated:

- that every property A is adequately represented by the binary variable V_A taking value 1 when A is instantiated and value 0 otherwise;
- that saying that a variable V causes a variable W cannot be saying anything else than saying that there exists (at least) one value v of V and one w value of W such that the property corresponding to v causes the property corresponding to w.

From these two remarks, one gets the idea that variable V_A causes variable V_B if and only if A causes B, or $not-A$ causes B, or A causes not $not-B$, or $not-A$ causes $not-B$. Letting the latter disjunction be "$(not-)A$ causes $(not-)B$", one comes to the idea that the comparison must be between BN characterization for "V_A causes V_B relative to a set of variables that suffices to describe empirical reality" – or, more simply, "V_A causes V_B" – and the analysis of "$(not-)A$ causes $(not-)B$" through probabilistic theories of causality.

B.2.3.2 Comparison of the analyses

The comparison turns on the classes of cases that are correctly treated by BN characterization on the one hand and probabilistic theories of causality on the other hand. Concerning probabilistic theories of causality, these classes are analogous to the classes of relations between properties that are correctly treated. As a consequence, the focus must be, here, on BN characterization.

1. BN characterization does not treat correctly spurious independencies between variables. Indeed, one easily shows that V_A and V_B are dependent if and only if A and B are dependent. Now it was shown (in subsection B.2.1) that B characterization implies that A and B are probabilistically dependent if A causes B. As a consequence, BN characterization fails when two properties are probabilistically independent and yet in cause-effect relation.

2. BN characterization does not treat completely spurious correlations due to common causes in interactive forks. This follows from the fact that V_A is dependent of V_B relative to $\mathbf{V}_C$ if and only if A is dependent of B relative to $\mathbf{C}$. In case C is an interactive cause common to A and B, BN characterization implies, sometimes wrongly, that V_A causes V_B.

From 1. and 2., one concludes that BN characterization faces more counter-examples than any of the probabilistic theories that are posterior to Suppes (1970).

3. Spurious correlations implied by a non interactive common cause are rightly identified as spurious by BN characterization. Indeed, if C is a cause common to A and B that makes them independent, then V_A and V_B are independent relative to the direct causes of V_A, be it V_C or an effect of V_C that is one of these direct causes.

As a consequence of 3., BN characterization does not face more counter-examples than Suppes' theory.

4. Spurious correlations that need conditionalization on *several* properties to be explained away do not raise problems for BN characterization. Indeed, the characterization requires that V_A and V_B be dependent relative to *all* the direct causes of V_A in order for V_A to cause V_B.

5. BN characterization gives a satisfactory treatment of spurious correlations between causes and effects. First, it does not appeal to the idea of one property being prior to another. Second, it implies neither that generic causation is symmetric, nor that it is asymmetric. Positively, conditionalization on the direct causes seems to adequately capture the facts that some but not all cause-effect relations are reciprocal.

6. As it turns on the identification of the properties that are in cause-effect relation, Simpson's paradox does not give rise to difficulties for the analysis of causality among variables. In other words, "$(not-)A$ causes $(not-)B$" is too coarse an object to be liable to the Simpson's paradox.

Together with 3., 4. to 6. implies that BN characterization faces strictly less counter-examples than Suppes' theory. In my history of probabilistic theories of causality, BN characterization occupies a place and this place is just after, but strictly after, Suppes' theory.

B.2.3.3 Consequences for causal inference

The question that finally arises is the following: does the situation of BN characterization in the area of probabilistic theories of causality explains why Bayes nets enable to infer probabilistic relations from probabilistic data?

As an answer, I claim first that the fact that spurious independencies are not taken into account is indeed important. More precisely, the fact that causation requires probabilistic dependence implies that direct causation requires probabilistic dependence. On the other hand, an extension of this is that A being a direct cause of B in a Bayes net over $\mathbf{V}$ is equivalent to probabilistic dependence relative to all subsets of $\mathbf{V}\backslash\{A, B\}$.[26] This is essential to the possibility of inferring generic direct causes in the way Bayes nets algorithms do. As a consequence, the fact that spurious independencies

[26]Spirtes et al. (1993) p. 82.

are not taken into account is essential to causal inference.

Second, it must be stressed that all this would be nothing is causation as it is characterized through Bayes nets were not relative to a set of variables. Indeed, this is what makes it possible to determine whether a set of variables screen off one variable from another.

In the end, BN characterization enabling to infer causes from probabilistic data is partly explained by both 1) the position of BN characterization of "V_A causes V_B" relative to probabilistic analyses of "$(not-)A$ causes $(not-)B$" and 2) the fact that BN characterization and probabilistic theories of causality do not have exactly the same objects. This concludes my enquiry concerning the criterion for causation that is used by Bayes nets causal inference. Correlatively, I now come to the methodological aspect, that is to the methodological features of Bayes nets causal inference.

B.3 Bayesian networks and causal inference

In the present section, I examine Bayes nets causal inference from a methodological point of view. Once again, the examination is justified both generically and specifically. Generically, it belongs to the analysis of the epistemological counterparts of probabilistic theories of causality. Specifically, it aims at determining whether Bayes nets allow to *induce* causes while causal inference is traditionally stuck to hypothetico-deduction.

As it may be already clear, my interest is mainly in the logic of causal inference, and hence in its principles. Procedures and implementation are evoked only secondary, in a subordinate way. Furthermore, as in section B.2, the methodology taken up is comparative. I set out first BN causal inference and second more traditional methods with which it competes. Third, I make the comparison. Fourth I explore the methodological recommendations that follow from the preceding analyses.

B.3.1 Bayes nets causal inference

Bayes nets causal inference develops out of algorithms. More precisely, those algorithms build the pattern that represent all the directed acyclic graphs relating to a given probability distributions along the lines of both the causal Markov condition (CMC) and the Faithfulness hypothesis. As they are the keystone of BN causal inference, those algorithms are set out first, and from this presentation I extract the principles of BN causal inference.

My presentation of BN causal inference algorithms focuses on algorithm PC by Spirtes, Glymour et Scheines. This means in particular that I am content with the simple case when sets of variables are causally sufficient. The steps of algorithm PC are described and it is explained that they do not suffice to compose a causal inference procedure. More exactly, using PC to infer causes from probabilistic data in general requires to first test (meaning, statistically test) for the probabilistic independencies between the considered variables.

Closer examination of the instructions composing PC reveals that algorithm PC identifies all and only the direct causal relations that follow from input probabilistic dependencies under Acyclicity (of direct causal graphs), the CMC, and Faithfulness. This leads me to distinguish between two notions of induction. The first one has to do with the relationship between the nature of premises and the nature of the conclusion, the second one with the logical relationship between the premisses and the conclusion. As far as the first one is concerned, Bayes nets causal inference is inductive: general knowledge is drawn from data about particular cases. In other words, BN

inferential strategy is inductive, meaning a-theoretical. As to the second one, Bayes nets causal inference is not inductive, but rather deductive: the conclusion is a necessary consequences of the premises – meaning here: of the treated probabilistic data.

B.3.2 Traditional causal inference

B.3.2.1 What should BN causal inference be compared with?

Obviously, to start with, one has to identify the more traditional causal inference methods with which BN causal inference should be compared. This is most naturally done by first identifying the kinds of phenomena to which BN causal inference is most suitable. As BN causal inference takes *observational* probabilistic data as premises, it is well-suited for phenomena that are studied by the social sciences.[27]

BN causal inference has a second remarkable feature: it is an inference to the causal *structure* over a set of variables. This is not always the case in the social sciences, where causation is often studied *locally*. As a consequence, it becomes more definite what BN causal inference should be compared with. Explicitly, BN causal inference methods should be compared with causal inference methods from structural equation modeling – or, more simply, causal modeling. As the section focuses on causally sufficient sets of variables, the methods that will interest me are from path analysis (PA) in the sense of Kline – which is more general that the original one.

It has been shown that BN causal inference should be compared with PA causal inference. What would be natural, then, would be to describe PA causal inference procedure, and to let PA causal inference principles emerge from this procedure – exactly as in the BN case. However, for various reasons, there is no such thing as *a* PA causal inference procedure. On the contrary, PA is can be characterized by specification of causal models, and the main feature of PA causal inference consists of its beginning by specifying a causal model. In other words, PA causal inference is best characterized at the levels of principles, as hypothetico-deductive. This leaves open the question of what is a PA causal inference procedure.

B.3.2.2 Traditional causal inference procedure

In the present paragraph, I set out the causal inference procedure that follow from both the principle of hypothetico-deduction and the tools of PA. This

[27] "Social sciences" is given its Anglo-Saxon meaning. So understood, they are made up of the studies which deal with man and in which statistical tools play an essential role.

procedure is ideal and is related to actual PA causal inference in the following way: the often noticed methodological weaknesses of actual PA procedures could be measured by the distance between these procedures and the one I define.

This procedure is made up of six steps: given a set of variables $\mathbf{V}$,and observational probabilistic data concerning $\mathbf{V}$,
Step A: specify an over-identified causal model over $\mathbf{V}$, say M.
Step B: estimate the parameters that are associated to the direct cause-effect relations represented in M.
Step C: Test M and possibly reject it.
Step D: Reiterate step A to C for causal models that are different from M.
Step E: Identify the one among the non-rejected of these models that best fits the data, say $M*$.
Step F: Identify the one among the models equivalent to $M*$ that most plausibly represents the causal structure over $\mathbf{V}$, say $MI_{\mathbf{V}}$.
$MI_{\mathbf{V}}$ is the output of the procedure.

What is exactly to be done at each moment and how this is done using PA tools deserve explication. Here I will be content to explain that I can envisage three kinds of tests for step C:

a. examine whether the signs and absolute values of estimated parameters are plausible. This deserves global as well as local consideration. From the local point of view, one must in particular ensure that each of the parameters is significantly different from zero;

b. calculate the differences between model-implied correlations and observed correlations and check that they are not too important[28];

c. calculate the over-identification restrictions – that is the differences between two estimates of the same parameter – and ensure that they are not significantly different from zero.

From the methodological point of view, it seems clear that the defined procedure is hypothetico-deductive. More precisely, as recommended by Popper himself, causal hypotheses are put to the test in two phases. First (step C), models are tested in isolation: one checks that estimation does not reveal an inconsistency. Second (step E and F), they are compared. The fact that E and F take place after C allows the comparison to turn only on serious candidate models, that have not been rejected after C. E and F can be considered a matter of inference to the best explanation.

[28] A rule of thumb is that they should be greater than .1.

B.3.2.3 Implementation

Finally, as far as implementation is concerned, the most important remark is that the importance of model specification results in PA procedures not being suitable for automation. On the contrary, BN causal inference being a-theoretical naturally leads to automatic implementation, through TETRAD packages in particular. It must be noticed that the LISREL packages include the possibility of automatic PA causal inference.

B.3.3 Comparison

B.3.3.1 Principles of causal inference

As far as principles are concerned, it has already been claimed that BN causal inference is deductive, meaning that the conclusion is BN inference is a necessary consequence of the premises. As a first moment of the comparison between BN causal inference and PA causal inference, one can show that PA causal inference is not deductive in this sense.

That the conclusion of PA causal inference is not a necessary consequence of the treated probabilistic data (here observed correlations) it takes as premises stems:

1. generically from its being hypothetico-deductive – which means, here, its following the lines traced in Popper (1934);

2. specifically from the way the Popperian recommendation to compare hypotheses can be followed using PA tools. More explicitly, remember that steps E and F of PA causal inference procedure can be described as inference to the best explanation. This is exactly the specific reason why PA causal inference is not deductive in the sense that BN causal inference is.

Deductivity, then, is a feature that is proper to BN causal inference.

This feature cannot be differentiated from BN causal inference strategy being a-theoretical. More precisely, deductivity of BN causal inference is necessary for BN causal inference strategy to be a-theoretical. On the other hand, it has already appeared that an a-theoretical strategy is suitable for, and in this sense corresponds to, automation of causal inference. As a consequence, I propose to isolate the effect of deductivity by comparing BN causal inference led by TETRAD to PA causal inference led by LISREL. Not surprisingly, the comparison is clearly in favor of TETRAD: deductivity is an asset of BN causal inference.

As already explained, deductivity concerns the principles of causal inference. More generally, all that has been said so far concerns the principles of causal inference, and not its validity. Now coming to whether causal inference is valid looks both interesting and important for my comparison. Indeed, it looks as if there were restrictions to the validity of causal inference both in the PA and in the BN cases, but these restrictions were very different in the PA case and in the BN one. The next two paragraphs are devoted to discuss these restrictions.

B.3.3.2 Restrictions to the validity of PA causal inference

Standardly discussed restrictions to the validity of PA causal inference are as follows:

1. in step B of PA causal inference procedure, parameters of specified models are not deduced, but only estimated from the data;

2. at the end of step C, the rejected models are not refuted properly speaking. Positively, in C, tests always deal with statistical hypotheses which can refuted only in the methodological sense that one decides that they do not make observed data probable enough.

In both cases, the consequences are new reasons why the conclusion of PA causal inference is not a logical consequence of its premises.

In both cases too, the problem consists in the premises of PA inference not being probabilistic correlations in the *population*, but only in a sample. Now this analysis reveals that the restrictions just discussed are not specific, in the end, to PA causal inference. More precisely, they admit of BN analogues, that are active when inference starts with correlations in a sample, rather than with probabilistic independencies in the population. The point is not new[29], yet it remains largely overlooked.

B.3.3.3 Restrictions to the validity of BN causal inference

Concerning BN causal inference, the most obvious restriction to its validity is that Acyclicity, the Causal Markov Condition and Faithfulness may be violated. The violations first seem to imply that (deductive) BN causal inference is available only for sets of variables that satisfy the three hypotheses. However a closer look reveals that things are in fact worse than that. Indeed, one cannot know in advance – that is, before the causal structure is known

[29]See Humphreys and Freedman (1996) p. 117.

– if the hypotheses are satisfied. As a consequence, BN causal inference can never be trusted even though it can give correct conclusions.

The question arising now is whether these restrictions to the validity of causal inference are specific to BN causal inference. As far as Acyclicity is concerned, the answer is clearly positive. Yet I claim that PA causal inference relies on a conception of the relationship between causality and probability that is at least very similar to the one that is conveyed by the CMC and Faithfulness together. On the one hand, in general one does not hypothesize (at step A of PA causal inference procedure) a cause-effect relation between two variables that are not correlated. Moreover, as already mentioned, if it cannot be excluded that a causal parameter of the estimated model is different from 0, then the model is rejected at step C, more precisely on the occasion of test a. On the other hand, test b. relies on the idea that probabilistic dependencies are exhausted by causal dependencies. More precisely, it assumed that two variables are in cause-effect relation if their probabilistic dependence is not explained by other cause-effect relations. In total, PA causal inference relies on a conception of causation as probabilistic dependence that does not disappear by conditionalization.

I have just explained that possible violations of the CMC and Faithfulness do not constitute a problem that is specific to BN causal inference. On the contrary, the status of these hypotheses and the consequences of their possible violations are different in the BN case and in the PA one. First, it has already appeared that the hypotheses are used locally in the context of PA causal inference. Quite differently, they are all there is to BN causal inference and hence are used right through it. Second, the consequences of possible violations of the hypotheses that were put into light in the BN case derive from a-theoricity of BN causal inference. As PA causal inference relies on the specification of models, satisfaction of the hypotheses can be tested for. More generally, possible violations of the CMC and Faithfulness do not entail that PA causal inference can never be trusted.

B.3.3.4 Conclusions

In this subsection, I have shown:
1. that the problem raised by the idea of refuting statistical hypotheses is not specific to PA causal inference, and that the problem raised by possible violations of the CMC and Faithfulness is not specific to BN causal inference. As a consequence, what is proper to BN causal inference is its deductivity;
2. that deductivity is an asset of inferences in general, and of BN causal inference in particular;
3. that deductivity has a-theoricity as a necessary condition;

4. that a-theoricity is precisely what makes the possible violations of Acyclicity, the CMC and Faithfulness have the consequence that BN causal inference can never be trusted.

Does this imply that Bayes nets, and in particular the powerful algorithms in the style of PC, cannot contribute to causal inference? This point is discussed in next subsection.

B.3.4 How Bayesian networks could contribute to causal inference?

As previous analyses make clear, contribution of BN algorithms to causal inference cannot take the form of causal inference being reduced to these algorithms. In other words, what I look for is a way BN algorithms could be integrated to PA causal inference procedure and contribute to it.

There already exists a proposition of this kind, more carefully formulated in Williamson (2002). The proposition is to use BN algorithms in order to formulate causal hypotheses – that is at step A of PA causal inference. It has the advantage to fill the gap of Popperian methodology as far as formulation of hypotheses (here, causal) is concerned. However, the gap is not completely filled, since BN algorithms output patterns rather than directed acyclic graphs. Resort to theoretical knowledge is not avoided in the end, and one then does not seem very clearly what the methodological profit exactly is. Moreover, Williamson does not take into account the fact that the causal Markov condition and Faithfulness may be violated. On the one hand, BN algorithms are used blindly; on the other hand, no methodological consequences are drawn from the fact that PA causal inference relies on hypotheses similar to the CMC and Faithfulness.

This leaves us with the possibility of using BN algorithms in order to test causal hypotheses. More precisely, the test would be as follows: for a specified causal model, if it cannot be excluded that Acyclicity, the CMC and Faithfulness would be satisfied if the model were correct, perform BN algorithm and check that the model under consideration is compatible with the output pattern. This is compatible with taking into account possible violations of Acyclicity, the CMC and Faithfulness only if:

1. the test is performed just after the model is specified – that is between step A and step B of PA causal inference procedure;

2. it is coupled with a specification of causal hypotheses independent from probabilistic correlations;

3. specified models found to be such that:

    - Faithfulness would be violated if they were true are *not* rejected on the occasion of test a. of C if some estimated parameter does not differ significantly from zero;
    - the CMC would be violated if they were true are rejected unless one has very firm theoretical justification for them. Indeed, parameter estimation relies on something like the CMC and, as a consequence, one cannot avoid using the CMC unless one renounces full stop to PA causal inference. In this sense, the status of the CMC in PA causal inference differs from the one of Faithfulness.

B.4 Indeterminism and the causal Markov condition

Sections B.2 and B.3 dealt with the way Bayes nets allow to infer causes – the question being tackled from the point of view of conceptual analysis in section B.2 and from the one of causal inference methodology in B.3. In both cases, it has revealed essential that Acyclicity, the CMC and Faithfulness are liable to violations.

The enquiry that is carried out in the present section is orthogonal to the ones of sections B.2 and B.3. Indeed, what is at stake is no more the consequences of the fact that Acyclicity, the CMC and Faithfulness sometimes are violated; rather it is exactly to determine when they are satisfied. More precisely, the present section takes part in the debate on the CMC and is a discussion of Steel (2005). In this paper Daniel Steel claims to establish that the CMC is satisfied by systems with jointly independent exogenous variables, be these systems deterministic or not. The claim is meaningful in relation with the classic result according to which deterministic systems with jointly independent variables satisfy the CMC: Steel claims that the determinism requirement can be removed. In order to examine Steel's claim I start with setting out the classic result.

B.4.1 Determinism and the causal Markov condition: the classic result

To begin with it must be explained what determinism is in the context of BN causal inference. I claim that it is before all a property of sets of variables: a set of variables is deterministic if of any variable that is endogenous in that set is functionally determined by that of its direct causes in the set. From this, one can defined the notion of a system being deterministic: a system is deterministic if any set of variables representing all the properties whose instantiation depends on the system, is itself deterministic. Thus defined, determinism of a system is quite different from determinism as an hypothesis about the world.

It is well-known and easily shown that a deterministic set of variables with jointly independent exogenous variables satisfies the causal Markov condition. By extension, one can say that any deterministic *system* with jointly independent exogenous variables satisfy the CMC.

B.4.2 Steel's result

Steel's result is established in the framework of what he calls "causal functional models":

Definition B.4.1 (Causal functional model) *Let* $\boldsymbol{X}= \{X_1, X_2, \ldots, X_n\}$ *and* $\boldsymbol{U}= \{U_1, U_2, \ldots U_k\}$ *be two sets of variables.*
A causal functional model over $(\boldsymbol{X}, \boldsymbol{U})$ *is* $(\boldsymbol{E},p)$ *where:*

- *p is a probability distribution over* $\boldsymbol{U}$*;*
- $\boldsymbol{E}$ *is a set of n equations, each of which expresses a variable* X_i *as a function* f_i *of* $((\boldsymbol{X} \cup \boldsymbol{U}) \setminus \{X_i\})$*;*
- *the equations in* $\boldsymbol{E}$ *are causal generalizations;*
- *for any* X_i*, the set of its direct causes in* $(\boldsymbol{X} \cup \boldsymbol{U})$ *is a subset of the set of variables whose it is a function according to* $\boldsymbol{E}$*.*

What Steels shows is essentially the following:

Theorem B.4.1 (Steel's result) *Let* $M = (\boldsymbol{E}, p)$ *a causal functional model over* $(\boldsymbol{X}, \boldsymbol{U})$*.*
If the variables in $\boldsymbol{U}$ *are jointly independent, then the set* $\boldsymbol{DC}_M$ *of the variables in* $\boldsymbol{X}$ *together with their direct causes in* $(\boldsymbol{X}, \boldsymbol{U})$ *satisfies the CMC.*

I have no problem with Steel's result. More explicitly, I put into question, neither its truth, nor the correction of the proof given by Steel.

The relationship between this result and indeterminism is that indeterministic systems can be represented through causal functional models. More precisely, it is always possible to use a variable of the **U**-type to represent the probabilistic way an indeterministic direct cause acts on one of its effects.

B.4.3 Steel's result and the classic result

The present subsection aims at assessing whether Steel has grounds for claiming to have established that the determinism requirement can be removed from the classic result. Now, he actually has grounds for claiming this only if the **U** variables of a causal functional model M are exogenous variables of the system represented by M.

It can be noticed that Steel hesitates between "error terms" and "exogenous variables"[30] when naming the **U** variables of causal functional models.

[30] Steel (2005) p. 5.

As a consequence, I first seek to determine whether they are error terms of exogenous variables. In order to do so, I come back to usual causal functional models and explain that, in this usual context:

- a variable is exogenous in a set of variables **V** if it happens not to have any cause in that set;
- the error term associated with variable V_i in **V** represents "in one heap"[31] all the influences that contribute to determine the value of V_i and yet is not represented by its direct causes in **V**. More generally, error terms enable functional representation.

With this distinction in mind, it appears that some of Steel's **U** variables are exogenous variables in the usual sense and that the others represent one kind of influences usually represented by error terms. More explicitly, the latter **U** variables represent the probabilistic way indeterministic causes act their effects. But no **U** variable represents omitted causes or measurement errors, though they are represented by usual error terms.

The consequence of this analysis is that Steel's presentation of his result is misleading. It is false that his result makes the determinism requirement in the classic result superfluous. More precisely still, this is true only if one calls the **U** variables "exogenous variables", which is at odds with usual use. My criticism does not concern Steel's result itself, but rather the way Steel advertises it. It is my contention that he should not pretend that his result is a direct contribution to the debate on the CMC as it existed before Steel (2005). Under usual terminology it remains false that the CMC is satisfied by any system with jointly independent exogenous variables, be it deterministic or not.

B.4.4 Steel's contribution to the debate on the causal Markov condition

Yet it seems to me that there is a sense in which Steel (2005) does contribute to the debate on the CMC as it usually stands. In order to show it, I first come back to Steel's causal functional models. My first claim, then, is that they do not constitute a coherent representation format: either one talks about the observable properties of a system, or one is interested in functionally representing it. In the first case, there is no room for the ones of Steel's **U** variables that represent the way probabilistic causes act on their effects; in

[31] Cartwright (1999) p. 9.

the second case, these variables will not do neither since they fail to represent *everything* that is needed for functional representation.

However, the variables in Steel's **U** that represent the way probabilistic causes act on their effect convey the suggestion of separating the various influences that are represented "in one heap" by usual errors. In other words, Steel's causal functional models suggest to use error terms in a realistic way. This leads me to define realistic causal functional models. In this new framework, one can formulate a condition (labeled "IC") on systems which is sufficient for the CMC. Now:

- if this condition is satisfied by a deterministic set of variables, then it is satisfied by any set of variables differing from the preceding only by certain causes acting in an indeterministic way;

- the converse is not true.

In this exact sense, determinism is more favorable than indeterminism for the causal Markov condition. If this result is considered as a contribution to the debate on the CMC, then so must be Steel's paper.

This conclusion needs to be qualified in two ways. First, it makes sense only in the context of systems being represented through causal functional models – be they Steel's ones, usual ones or realistic ones. That this format of representation is appropriate is assumed by the proponents of the debate of the CMC, and consequently by me taking part here in this debate. Still the assumption could be criticized and one should wonder whether many systems of interest can be represented by a causal functional model. Second, the conclusion of the present section is far from being the last word in the discussion concerning the satisfaction of the CMC, let alone the debate on the extension of the domain inside which BN causal inference algorithms give correct results. Moreover, it seems clear that I will not close either of these debates. Correlatively, I now put an end to my enquiry on the epistemological correlates of probabilistic theories of generic causation, and I turn to the probabilistic analysis of singular causation.

B.5 The propensity theory of probability

In the second part of my thesis, which is now beginning, I come to conceptual analysis of singular causation. More precisely, my concern will be in the relationship between actual causation and probability. As already explained, probability in this context cannot be interpreted by in terms of propensity. Indeed, the propensity theory is the only interpretation of probabilities that can make sense of the notion of singular physical probabilities.

In the section now beginning, I set out the propensity theory of probability. The presentation deals essentially with Popper's propensity theory and is intended to have a large scope. More precisely, I envisage the propensity theory not only as a theory of probability, but also as a philosophical theory, or rather as a theory that cannot be separated from ontological, epistemological and even metaphysical options. These are put into light in the second subsection, the first one being devoted to set out the propensity theory.

B.5.1 Presentation of the propensity theory

As noticed in particular by Gillies[32], there exist many different propensity theories of probability. Moreover, this diversity stems precisely from the relative indetermination of Popper's position. As a consequence, my strategy is as follows: first giving a minimal characterization of the propensity theory – that is a characterization of the theoretical basis out of which the various propensity theories develop –, and then presenting in more detail the particular version in which I will be interested. In the third paragraph, I tackle the question of whether the propensity theory can be considered an interpretation of the probability calculus.

B.5.1.1 Minimal characterization of the propensity theory

Even giving a minimal characterization as the propensity theory as it is introduced by Popper is not straightforward. Indeed, the very notion of propensity is plurivocal. This plurivocity leads me to proceed as follows: I identify three noticeably different ways to characterize propensities, discuss them each individually and finally extract from this discussion a characterization of propensities and of the propensity theory that take all three into account.

First, propensities may be characterized as properties of sets of physical conditions. This is the consequence of the idea that the probability of a singular event is determined by the set of conditions that may produce it. As

[32]Gillies (2000a) p. 113.

a consequence, that probability is a property of this set of physical conditions, a property that is dispositional in the sense that it manifests itself only when the event occurs. This characterization of propensities has three main features: 1) it makes it very uneasy to distinguish between probability and propensity, 2) propensities are dispositional properties of a particular kind: there does not exist a set of conditions that makes it necessary that it gets manifested, 3) one should define more precisely the kind of sets of physical conditions of which propensities are properties.

Second, Popper sometimes characterizes propensities as metaphysical entities of the same kind as Newtonian forces. Here "metaphysical" essentially means: having a physical reality and yet being unobservable. This characterization suffers from the indecisiveness of Popper concerning the exact relationship between propensities and forces.

Third, propensities may be presented as possibilities. This is the case in Popper (1990). More precisely, Popper talks about "weighted possibilities". The matter is that the phrase is not fully explained. Moreover, the idea that propensities are possibilities is denied in other texts, for instance in Popper (1983).[33] In this text, propensities seem to be *measures* of possibilities, which makes it once again difficult to distinguish between probabilities and propensities.

Starting with this, I produce what I claim is the simplest among the descriptions of propensities that account for all three families of intuitions. The description starts with the idea that propensities are attached to, and hence are properties of, sets of physical conditions. Now propensities are metaphysical in the sense that they are unobservable and yet have physical reality. This reality entirely consists in their action towards the production of possible events. The occurrence of such an event manifests the existence of a propensity that tended to produce it. In this sense, propensities are dispositional properties. Propensities vary in intensity and their intensity is measured by probabilities. As a consequence, probabilities too are attached to sets of physical conditions. This picture makes propensities causal in two non independent ways: first propensities as dispositions are good candidates to count as causes of their manifestation, second propensities are metaphysical entities able to produce events and hence look very much like causes conceived in a realist way.

The picture I have just drawn is claimed to be the simplest way to arrange the various aspects of Popper's characterization of propensities. It is also minimal in the sense that it is the basis common to Popperian propensity theories. Out of this basis, various theories emerge. I have to come here

[33]Popper (1983) p. 371.

into this detail. More precisely, I come to the distinction between long-run propensity theories and single-case propensity theories, and I defend a single-case position.

B.5.1.2 In favor of single-case propensity theories

What I have to do first here is to explain what distinguishes single-case propensity theories from long-run propensity theories. Basically, the difference is about what propensities tend to realize: singular events on the one hand, relative frequencies on the other hand. Now, Gillies[34] explains that two other differences are subordinate to the basic one. First, long-run propensity theories go with the requirement that the sets of physical conditions to which propensities are attached be repeatable. On the other hand, most proponents of single-case propensities with Popper's attaching them to the "global physical situation"[35] at one given moment – which clearly excludes repeatability. Second, it follows that probabilistic hypotheses may be tested through frequencies in the long-run case, but not in the single-case one.

That they are not compatible with the testability of probabilistic hypotheses does not make long-run propensity theories unacceptable. More precisely, as already noticed by Gillies[36], it does not make them unacceptable as metaphysical theories. By extension, it does not make them unacceptable in the context of the analysis of the relationship between probability and actual causation. Moreover I claim that probabilistic hypotheses under a single-case propensity theory can often be tested through frequencies *in practice*. On the one hand, and as suggested by Popper, a particular set of physical conditions is often sufficiently characterized by repeatable properties. On the other hand, the concept of methodological refutation makes it possible to use observed frequencies in order to test probabilistic hypotheses. As a whole, it is my contention that the point of testability does not invalidate the idea of single-case propensity theory.

I have still to answer objections to the very notion of a singular objective probability. Two objections deserve particular attention: the one that is developed by Gillies[37] and the one developed by Kyburg[38] According to Gillies, probabilities of singular events cannot be objective since their assignation involve many subjective elements. According to Kyburg, our uses of the notion of objective probabilities do not call for *singular* objective proba-

[34]In Gillies (2000a).
[35]Popper (1990) p. 39.
[36]Gillies (2000b) p. 824.
[37]Gillies (2000b) pp. 813–817.
[38]Kyburg (2002).

bilities: relative frequencies do the job. What is central in both cases is the epistemic point of view. More explicitly, objective probabilities of singular events are envisaged only to the extent that they are objects of knowledge. Correlatively, the arguments are not about the very notion of physical probabilities of singular events that single-case propensities theories convey. As a consequence, I conclude that they do not invalidate single-case propensity theories of probabilities.

Up to now, I have answered the main criticisms that are addressed to single-case propensity theories. In the end of the paragraph, I give positive reasons to prefer them to long-case propensity theories of probability:
1. I point out the fact that, historically, the *raison d'être* of Popper's propensity theory is its providing an objective interpretation for singular probabilities. More precisely, the propensity theory emerged as a consequence of the frequency interpretation giving only an ersatz of what would be a singular notion of probability.
2. I explain that the possibility of conceiving the propensity theory as emerged from frequentism through emphasis put on the physical conditions that produce aleatory events. More explicitly, I explain that this constitutes a historical argument in favor of long-run propensity theories, but that this argument is not analogous to the one previously developed in favor of single-case propensity theories. The first argument led to the conclusion that long-term theories lack an essential feature of propensity theory. On the contrary, single-case theories do convey the emphasis on the physical conditions out of which aleatory events are produced.
3. Finally, I raise two objections against long-run propensity theories. According to the first one, they do not differ significantly from frequency theories. More precisely, it is not very clear how the differ from well-understood frequentism.[39] Second, I do not understand very well what it can mean for a propensity tending to produce actual frequencies. If propensities really are metaphysical entities, then the notion of a long-run propensity is hardly intelligible.

B.5.1.3 Is the propensity theory an interpretation of probability?

What remains to be done is determining whether a single-case extension of the Popperian minimal basis for propensity theories can be considered an interpretation of probability calculus. In order to do this, I come back to Kolmogorov's theory of probability (denumerable additivity being excluded from the discussion), and I examine what it can mean to be an interpretation

[39] Consider for example Kolmogorov's position.

of this theory. It appears in particular that it cannot be understood strictly in the model–theory sense of "interpretation". Indeed, interpretations of probability convey a proposition as to the nature of probabilized objects, whereas Kolmogorov's axioms implies that probabilities are probabilities of *sets*. Positively, interpreting the probability theory can be considered as:

- giving a set $\mathbf{o}$ closed by union and complementation,
- identifying an element w in $\mathbf{o}$,
- defining a function p from $\mathbf{o}$ to $[0; 1]$

which are such that $(w, \mathbf{o}, p)$ satisfy Kolmogorov's axioms.

Under this explication, there is a fundamental obstacle to the propensity theory being an interpretation of probability. More precisely, there is a difficulty in *showing* that the propensity theory is an interpretation of probability. Indeed, contrary to the frequentist and subjectivist interpretations, the propensity theory does not convey a theory of how probabilities should be evaluated. As a consequence, there is no empirical answer to the question of whether the propensity theory is indeed an interpretation of probabilities.

In spite of the impossibility to answer the question empirically, one can give arguments on favor of the idea the the propensity theory provides an interpretation of probabilities. To begin with it must be noticed that the proposed notion of interpretation of Kolmogorov's theory clings to model–theory and as such to the nowadays dated idea of probability theory as an extension of logic. Positively, it seems nowadays clear that admissibility is far from being all there is to interpreting probabilities. One other possible criterion is applicability. Now, with regard to this criterion, the propensity theory has non negligible assets. In particular, it enables to distinguish probabilities from frequencies, it can account for the fact that more probable events occur more frequently, it accounts for the use of probabilities in quantum theory.[40] To finish with there is Lewis' argument in favor of the propensities being probabilities since they are equal to degrees of belief. It relies heavily on the validity of the Principal Principle, which precisely can be criticized. As a consequence I do not consider the argument conclusive. Still the idea that there must be a connection between propensities and degrees of belied suggest that propensities must have probabilistic features. As a whole, I have shown that the hypothesis according to which probability theory is an interpretation of probability may be accepted, and I will actually accept it.

[40]On these points, see in particular Hájek (2007).

B.5.2 The philosophical dimension of the propensity theory

In this subsection, I examine the philosophical correlates of single-case propensity theories stemming from Popper's characterization of propensities. The examination is mainly comparative: the philosophical correlates of propensity theories are compared with the ones of other probability theories. It deals with ontology, epistemology and finally metaphysics.

B.5.2.1 Propensity ontology

As far as ontology is concerned, the analysis originates in the distinction between two traditions. These two traditions diverge concerning the existence of things that are different from full or empty space-time points. Clearly, propensity theories commit to the thesis that such things exist. What I examine, then, is whether and how this makes a difference with the frequentist and the subjectivist theories. In both cases, the discussion is largely about operationalism. Indeed, although an epistemological option, operationalism clearly has reductionist ontological consequence.

As far as frequency theories are concerned, I am interested in the more appealing version, that is the limiting-frequency theory. I claim that this theory is not operationalist: as noticed by Gillies[41], limiting frequencies do not measure probabilities in the sense that the column of mercury measures temperature. Moreover, the discussion suggests that the limiting-frequency theory leads to assume the existence of infinite series. Now these series are not reducible to empirical series. To this extent, frequentism leads to accept the existence of something different from space-time points. The difference with propensity theory is in the nature of what is thus accepted: natural entities in the propensity case, ideal objects in the frequentist one.

As far as subjectivism is concerned, it has as a part the theory of measurement of probabilities in bet situations. Due to this, the subjectivist theory is operationalist and thus does not belong to the same ontological tradition as propensity and frequency theories. This being established, I go a little further. Indeed, I first explain that subjectivism concerning probabilities can be considered as an ontological project: understanding probabilities without reference to any entities but space-time points. Moreover, it is shown through an analysis of de Finetti (1937) this project for understanding probabilities naturally extends to a regularity conception of laws of nature, and then to a general ontology in the type of Hume's. By contrast, a feature of propensity ontology appears to be ontological realism, that is the thesis according to

[41] Gillies (2000a) p. 100.

which there exists a physical reality existing independently from individuals able to perceive and think.[42].

B.5.2.2 Propensity epistemology

The ontological distinction between propensity realism and subjectivist anti-realism has important consequences in the field of epistemology. Basically, ontological realism cannot be separated from the thesis that it is possible for what is held true to diverge from what is true. As a consequence, minimizing this difference becomes an aim, in general and concerning probabilities in particular. On the contrary, probability assignations under radical subjectivism are not even likely to be true or false. This important difference between proponents of propensities and subjectivists can be shown to stem from from their drawing different consequences from the law of large numbers. For de Finetti, the law implies that probabilistic assignations can never be tested, whereas Popper introduces methodological refutation.

I have shown that the ontological distinction between propensity realism and subjectivist anti-realism extends to epistemology. I explore at least three consequences of this fundamental epistemological divergence:

- concerning truth, the proponent of propensities conceives of it as correspondence. On the contrary, subjectivism is compatible at most with its characterization as coherence;
- concerning rationality, it is a matter of each probability assignation under propensity theory. On the contrary, subjectivism leads to conceive of it as a structural matter;
- concerning inter-subjectivity, the propensity theorists sees it as a sign of the truth of the proposition on which various individuals agree. On the contrary, inter-subjectivity is an empirical phenomenon that it is essential and difficult for subjectivists to explain.

B.5.2.3 Propensity metaphysics

Propensities not only commit to the idea that there exist entities not reducible to space-time occupation. Indeed, propensities are active entities: they tend to make events happen. As a consequence, the propensity theory has metaphysical correlates. More precisely, propensities looking like Aristotelian potentialities, they seem to commit to a Aristotelian metaphysics. This hypothesis is discussed and qualified.

[42]Notice that this terminology is different from the one used by Popper in Popper (1983).

First, I compare propensities with potentialities. This is often done by Popper, who insists on the fact that potentialities are attached to individual things whereas propensities are properties of sets of physical conditions. I bring out three other differences between propensities and potentialities:

- potentialities are never considered as autonomous entities, whereas it is the case of propensities in the Popperian context to which I have stuck in my initial presentation;
- potentialities are not liable to quantification;
- Aristotelian potentialities are introduced in the context of the description of change, which remains to be explained through the theory of causes. On the contrary, propensities can be taken to explain observable frequencies and the fact that they are stable and tend to converge. Propensities do not call for an explanatory theory.

It is my contention that these differences make it all the more remarkable the convergence of the metaphysical correlates of potentialities and propensities. More precisely, I propose to consider that both commit to what I call "change metaphysics".[43] This can be characterized by coming back to possibility. The notion is central both in Aristotle and in Popper. On the one hand, a being is potentially everything that in can be; on the other hand, propensities tend to realize possible events. Moreover, both potentialities and propensities lead first to attach some kind of reality to possibility. Second, they both lead to the idea of actual existence as this kind of reality having utmost degree. Now, change metaphysics as characterized by these two theses is correlative the idea that change is not necessary and yet takes place inside the limits that are fixed by an initial configuration. Correlatively, the world is neither chaotic nor completely determined.

Aristotle does not articulate these two features. On the contrary, Popper develops the idea of scientific determinism as an approximation of indeterminism in the world.[44] Now the notion of scientific determinism is quite new – it appeared with Laplace. As a consequence, the difference between Aristotle and Popper concerning the analysis of the relationship between the two essential features of change can be thought in terms of Popper's metaphysics being a *modern* change metaphysics. I identify two other differences between propensity metaphysics and Aristotle's theory of change and claim that both can be thought along the same lines. On the one hand, Popper

[43]The phrase is introduced in Bouveresse (1981) (p. 128) concerning Popper's metaphysics.

[44]Popper (1982a) last part.

envisages change metaphysics as a mean for metaphysical indeterminism and with it metaphysical freedom[45] – an eminently modern notion. On the other hand, the fact that propensities are attached to physical conditions rather than to individual beings is essential to the novelty being possible. On the contrary, Aristotle's theory of change does not make any room for new things appearing. Ancient thought is about a closed world.

This puts an end to my analysis of metaphysics propensity. The analysis has made clear that the propensity theory of probability commit to a modern change metaphysics. The fact that it commits to a metaphysics is undoubtedly the most original of the philosophical correlates of the theory. Concerning these correlates more generally, it must be underscored that my whole analysis relies on the characterization proposed in the first subsection. But this characterization is based on Popper's account, which I extended by a defense of single-case propensities. As a consequence, one may wonder whether some aspects of the analysis in the second section specifically concern Popper's propensity theory. To this question, I answer that this is the case of the ontological analysis. Indeed, the idea that propensities exist physically and autonomously is far from being accepted by all propensity proponents. Apart from those in paragraph B.5.2.1, the conclusions of the second subsection can be taken to hold for propensity theories in general. It is also the case of the analyses that are carried out in the section to come.

[45]Popper (1982b) p. 78.

B.6 Causality and the propensity theory of causality

Everything that was said in the last section concerns the propensity theory as a theory of *absolute* probabilities. But probabilistic theories of causality that I am interested in rely on the idea to analyze causality through *conditional* probabilities. As a consequence, I now turn to conditional probabilities under a propensity interpretation. Then I am confronted with Humphreys' paradox, that is roughly to the idea that there cannot be a propensity interpretation of probabilities. The paradox is more precisely presented in the first subsection. Second, I analyze the disagreement between Paul Humphreys and one of its opponents in this debate: Christopher McCurdy. Third I try to build a propensity interpretation of conditional probabilities. My proposition is discussed in subsection B.6.4. Fifth and finally I can come to the question of the relationship between actual causation and probabilities given a propensity interpretation.

B.6.1 Humphreys' paradox

Humphreys' paradox has several formulations, some of them being formal and others being informal. To begin with, I set out an informal version which clings to the presentation of the propensity theory of probability that was given in section B.5.

B.6.1.1 Informal version

This informal version relies on the idea that there is a sense in which probabilities given a propensity interpretation are all conditional, since they are all relative to a set of physical conditions. More precisely, it seems that the propensity theory leads to consider probabilities as conditional probabilities whose conditioning elements are sets of physical conditions and whose conditioned elements are events that those sets may produce. Under this conception, the relationship between the conditioning element of a probability and its conditioned element looks very much like a cause-effect relations. As a consequence, it is asymmetrical. On the contrary, probability calculus is symmetrical relative to conditionalization. Then, the argument goes, the probability theory cannot be an interpretation of conditional probabilities.

This first version of Humphrey's paradox reveals irrelevant in the light of Gillies' distinction between fundamental conditional probabilities and event-

conditional probabilities.[46] More precisely, the fact that probabilities are relative in the context of the propensity theory does not entail that they are event-conditional probabilities. Rather they are fundamental conditional probabilities, with the consequence that the argument I have just presented is of no relevance relative to the question of whether causation, which is a relation between events, can be analyzed in terms of conditional probabilities. Still Humphreys' paradox as it is presented in Humphreys (1985) deals with event-conditional probabilities. As a consequence, I come to this paper. More precisely, my interest will in the formal argument that is developed in this paper.

B.6.1.2 Humphreys' paradox for event-condition probabilities

Humphreys' paradox appears in relation to the question of the evaluation of inverse conditional probabilities, that is probabilities whose conditioning event is posterior to their conditioned event. Humphreys claims that, under a propensity interpretation of probability, such a probability – say $p(A_{t2}|C_{t3})$ – should be evaluated following principle (CI): $p(A_{t2}|C_{t3}) = (p(A_{t2}|\overline{C_{t3}}) =)p(A_{t2})$. The justification is as follows: $t2$ being posterior to $t1$, the propensity for A_{t2} cannot be affected by the occurrence of C_{t3} and thus $p(A_{t2}$ is not modified by conditionalization on C_{t3}.

The paradox consists in the fact that principle (CI) is in formal contradiction with theorems that are essential to probability calculus as a calculus of both absolute and conditional probabilities. Humphreys concludes that "propensities cannot be probabilities".

Obviously there is only one way out of Humphreys' paradox: showing that inverse conditional probabilities should not be evaluated following (CI). Indeed, at least two rival principles have been proposed for the evaluation of inverse conditional probabilities in a propensity context:

- principle (ZI): if C_{t3} is posterior to A_{t2}, then $p(A_{t2}|C_{t3}) = 0$;
- principle (FP): if C_{t3} is posterior to A_{t2}, then $p(A_{t2}|C_{t3}) = 0 or 1$, depending on whether A_{t2} occurs or not.

Both principles stem from conceptions of what is the propensity account of conditional probabilities. Relying on this, Humphreys' paradox can be generalized. More precisely, Humphreys (2004) shows that 1) all existing conceptions of how the propensity theory accounts for conditional probabilities lead to either (CI), or (ZI), or (FP) and 2) both (ZI) and (FP) lead to difficulties that are exactly analogous to the one raised by (CI).

[46] Gillies (2000a) p. 132.

Among the existing propensity accounts for conditional probabilities, those that Humphreys calls "co-production interpretations" hold my attention. They rely on the idea that conditional propensities are localized in the initial structural conditions[47] and I claim that Humphreys does not treat them on the same footing as the other propositions he considers. First, he discusses more at greater length than their rivals. Second, proponents of co-production interpretations do not advocate any of the three principles identified by Humphreys, and Humphreys has to show that they are committed to one of them (in fact principle (CI)). If Humphreys' argument is not valid, then co-production interpretations may still convey a solution to the paradox. Third, it seems to me that Humphreys' concern about co-production interpretations consists at least in their content as in the way they lead to evaluate inverse conditional probabilities. More explicitly, Humphreys considers that co-production interpretations are problematic as propensity interpretations of conditional probabilities.[48] This suggests that Humphreys' dissatisfaction with co-production interpretations has to do directly with the question of the propensity interpretation of conditional probabilities. As a consequence, analyzing the disagreement between Humphreys and proponents of co-production interpretations should shed light on the origin of the paradox and possibly reveal how it may be solved. More precisely, I will focus on the disagreement between Humphreys and McCurdy, who both defends a standard co-production position and proposes an extensive discussion of Humphreys (1985)[49].

B.6.2 The disagreement between Humphreys and McCurdy

The disagreement between Humphreys and McCurdy first concerns the example that is introduced by Humphreys in order to introduce principle (CI). Concerning this example, McCurdy's criticism deals with Humphreys' justification of (CI) through consideration of alterations of the envisaged system. According to McCurdy the alterations concern what happens *after* A_{t2}, and then do not have any consequences on $p(A_{t2})$ – in particular they do not show that $p(A_{t2}|C_{t3}) = p(A_{t2})$. It is claimed in Humphreys (2004) that McCurdy is misled by some features of the system under consideration. Still, I argue, Humphreys does not give any new argument in favor of (CI), but rather reasserts his position. Moreover, his arguments against McCurdy are

[47] Humphreys (2004) p. 671.
[48] See Humphreys (2004) pp. 673 and 677.
[49] McCurdy (1996).

not in a position to convince that McCurdy is wrong. As a whole, it is my contention that the disagreement is quasi-unchanged by Humphreys (2004), and that precisely this reveals that the disagreement is not only about the evaluation of one particular inverse conditional probability.

First, the disagreement extends to the evaluation of inverse conditional probabilities in general. As already stated, Humphreys considers that they should always be evaluated in accordance with (CI). On the contrary McCurdy does not seem to accept any *principle*, but rather considers that one always has to come back to the initial set of physical conditions in order to evaluation probabilities, be they absolute or conditional.

Second, and as already suggested, the disagreement is about the propensity interpretation of conditional probabilities. Concerning Humphreys, the way he explains (CI) reveals that the propensity theory leads to consider that $p(A_{t2}|C_{t3})$ measures the propensity to produce A_{t2} *as it is possibly modified by the occurrence of* C_{t3}. By contrast, McCurdy does not consider that the difference between $p(A_{t2}|C_{t3})$ and $p(A_{t2})$ corresponds to the possible effect of the occurrence of C_{t3} on the propensity to produce A_{t2}. Rather it corresponds to what the occurrence of C_{t3} logically implies concerning the initial sets of physical conditions. In other words, conditionalization is interpreted as a re-specification, depending on the conditioning event, of the set of initial conditions.

Third, I claim that Humphreys and McCurdy disagree about the status of the propensity theory. To make it clear, I come back to Humphreys' dissatisfaction with McCurdy's co-production position. What Humphreys considers problematic is the fact this position fails to make conditionalization "a material relation between concrete events"[50]. According to Humphreys, this leads McCurdy to overlook what makes the propensity theory appealing, that is the fact that it puts the stress on physical dispositions. I maintain that this implies that Humphreys considers that the main contribution of the propensity theory to the philosophy of probability is the light shed on some probabilistic objects (indeterministic dispositions). The propensity theory is first the theory of these objects, which so to say "happen" to be probabilistic. Correlatively, what would be the propensity theory of conditional probabilities is analytically contained in the propensity theory of absolute probabilities. More explicitly, the propensity theory of conditional probabilities is the theory of second-order indeterministic dispositions – "conditional propensities" – that happen not to be conditional probabilities. By contrast, McCurdy sticks to the idea of interpreting probabilities. In conformity with what Lewis' triviality results suggest, this leads him to conceive of condition-

[50]Humphreys (2004) p. 675.

alizing as redefining the probability function.

In the end, I have shown that Humphreys' paradox cannot be separated from the idea that the propensity theory of absolute probabilities analytically contains a theory of conditional probabilities. But the variety of propensity interpretations of conditional probabilities shows that this is not the case. As a consequence, the question of the propensity interpretation of conditional probabilities can be considered unsettled. More positively, it remains logically possible to propose a propensity interpretation of conditional probabilities that would avoid the difficulties highlighted by Humphreys. This is what I try to do now.

B.6.3 Proposition for a propensity interpretation of conditional probabilities

B.6.3.1 What it is to interpret probabilities

In order to propose a propensity interpretation of probability, I examine what it is to interpret probabilities in general, and conditional probabilities in particular. As already mentioned, Lewis' triviality results suggest that conditionalizing is redefining the probability function (rather than substituting a complex argument for a simple one). As a consequence, I start with the idea an interpretation of probability calculus as a theory of absolute and conditional probabilities consists of:

1. an interpretation of absolute probabilities;

2. an analysis of the way conditionalizing redefines a given probability function.

This analysis is shown to being coherent with the idea that frequentism and subjectivism are interpretations of probability calculus. Moreover, examining these theories leads to distinguish between what determines probability functions – that is, in Gillies' words for the propensity theory, the fundamental conditioning element – and what is likely to be conditioned on – that is conditioning events. Armed with this distinction, I claim that conditionalization can be considered as the redefinition of a fundamental conditioning element by a conditioning event. As a consequence, interpreting conditionalization is explaining how a given conditioning event redefines a fundamental conditioning element. In other words, conditionalization has to be interpreted through a function which associates a new fundamental conditioning element with the couple composed of a fundamental conditioning

element and a conditioning event. It is explained what this function is in the frequentist and subjectivist contexts.

In total, an interpretation of probability calculus is composed of:

1. an interpretation of absolute probabilities. This must in particular specify:
   (a) the nature of fundamental conditioning elements;
   (b) the nature of the arguments of probability functions;
2. a definition of the function meant to interpret conditionalization.

B.6.3.2 McCurdy's interpretation

First it must be noticed that McCurdy and Humphreys seem to agree on the idea that the propensity theory (as I set it out in section B.5) is:

1p. an interpretation of absolute probabilities, under which:
   (a) fundamental conditioning elements are sets of physical conditions;
   (b) arguments of probability functions are singular events liable to be produced by those sets of physical conditions.

As a consequence, a propensity interpretation of conditionalization is a function which associates a new set of physical conditions with a set of physical conditions together with a singular event.

What I do here is bringing into light the way McCurdy would define this function. An examination of McCurdy (1996) leads me to the conclusion that McCurdy would define this function as follows:

$$c_{mc} \quad : \quad (p_1 \wedge p_2 \wedge \ldots \wedge p_n, p) \quad \longmapsto \quad p_1 \wedge \ldots \wedge p_n \wedge p.$$

where $(p_1 \wedge p_2 \wedge \ldots \wedge p_n$ describes the initial set of physical conditions and p is the proposition that the conditioning event occurs.

Now I see a fundamental difficulty with this interpreting function: its arguments are not sets of physical conditions together with singular events, but rather *descriptions* of sets of physical conditions together with *description* of singular events. This means that the proposed interpretation is not exactly compatible with the propensity interpretation of absolute probabilities. Above all, it is not easily translated into a proposition exactly compatible with the propensity interpretation of absolute probabilities. Indeed it is not clear what the conjunction of a set of physical conditions together with a singular event may be.

B.6.3.3 Construction of a proposition

In this paragraph I construct a function that I propose as a propensity interpretation of conditionalization. As just explained, this function will associate a new set of physical conditions with a set of physical conditions together with a singular event. But one can imagine many functions of this type. As a consequence, I have to impose requirements on the function I am looking for. To start with, I notice that it will not be an interpretation of conditionalization unless it accounts for the following property:

> (PC) For any probability function P and any argument A of P such that $P(A) \neq 0$, $P(A|A) = 1$.

In other words, the interpretation c_p that will be proposed must be such that:

> For any set B_{t1} of physical conditions and any singular event such that $Pr_{B_{t1}}(A_{t2}) \neq 0$, $Pr_{c_p(B_{t1},A_{t2})}(A_{t2}) = 1$.

This very first requirement implies that conditionalization cannot be interpreted as a "temporal jump"[51]. Indeed consider the set of physical conditions B_{t1} – that is the physical system B at time $t1$, and a singular event A_{t2} that this system may produce. Then there may not be a time when B_{t1} gives probability one to A_{t2}. More generally, it appears that it will not be sufficient to make time vary: systems themselves will have to change with conditionalization.

A natural idea, then, is to interpret conditionalization as the function associating with (B_{t1}, A_{t2} the system which is the most similar to B_{t1} among those in which A_{t2} occurs. The problem with this proposition is that it does not seem to be compatible with indeterminism. More exactly, in case of indeterminism, it is not always true that the system B' which is the most similar to B among those in which A_{t2} occurs is such that $Pr_{B'_{t1}}(A_{t2}) = 1$.

This suggests that not only the system, but also the time has to vary with conditionalization. In other words, it must associate with (B_{t1}, A_{t2}, the system B'_{t3} where $t3$ is a time such that $Pr_{B'_{t3}}(A_{t2}) = 1$. Such a $t3$ exists A_{t2} occurs in B'. However it is not clear how $t3$ may be generically defined at $t1$. Moreover, under indeterminism it is not always determined at $t1$ whether A_{t2} occurs. To finish with, the idea that both both time and systems vary with conditionalization is not appealing. Positively, I am led to the idea of considering not only actual, but possible systems.

Explicitly, my proposition is as follows: interpreting conditionalization through the function associating with (B_{t1}, A_{t2} the set of physical conditions

[51] The term is coined by Humphreys: Humphreys (2004) p. 672.

$B*_{t1}$ where $B*$ if the most similar to B among the systems which give A_{t2} probability 1 *at time* $t1$. In favor of this proposition, I argue that:

- it satisfies (PC) by construction;
- it is minimal in several senses: it emerges as necessary from preceding discussion, it makes only systems vary under conditionalization, the variation is to the most similar;
- what seems its most problematic feature – namely the appeal to *possible* systems – cannot count against it. First, whoever tries to give a propensity interpretation of conditionalization has already accepted that probabilities refer to unobservable objects. Second, the objection does not resist a widening of theoretical horizon. More precisely, I cannot see any substantial ontological difference between accepting possible worlds and accepting possible systems. As a consequence, considering the problem of interpreting conditional probabilities not in isolation, but rather as part of a class to which analyzing counterfactuals belongs, implies that the notion of possible systems cannot be too high a price to pay in order for the problem to get solved. An analogous reasoning leads to accept the idea of a function associating with each system, the system of a given class to which it is most similar.

B.6.4 Discussion of the proposition

It remains to be determined wether the proposed interpretation of conditionalization solves Humphreys' paradox. In other words, it remains to be determined whether the interpretation I propose for conditionalization accounts for the formal properties of Bayesian conditionalization. This, obviously, is a difficult question. As a consequence of its difficulty, the question is first stated in negative terms. More exactly, I first examine whether probabilities under this interpretation are not probabilities of conditionals rather than conditional probabilities.

B.6.4.1 Probabilities of conditionals?

There exist at least two arguments in favor of this hypothesis:
1. the way I propose to interpret conditionalization leads to consider that $P(B|A)$ is the probability of B in the system which is the most similar to the system under consideration among those that give A proba-

bility 1.[52] This sounds very much like the probability of the conditional $[P(\text{“}Aoccurs\text{”}) = 1] > \text{”}Boccurs\text{”}$ if this conditional is analyzed as suggested by Stalnaker. The distinction between possible systems and possible worlds is shown to be inessential here.
2. it is shown in Lewis (1976) that the probability $P(A > B)$ of the conditional $A > B$ is the probability of B for the probability function P_A^i that stems from P through imaging on A. It is shown in Walliser and Zwirn (2002) that the distinction between Bayesian conditionalization and imaging corresponds to the distinction between revising and updating. But the difference between revising and updating is the difference between taking into account of new information concerning an unchanged world and taking into account changes in the world.[53] Now, the propensity theory is interested in objective features of the physical world. Moreover, the way I propose to interpret conditionalization is indeed in terms of changes in the system under consideration. As a consequence, it seems that what I propose is an interpretation of imaging rather than an interpretation of Bayesian conditionalization.

However convincing they may see, these two arguments are not valid. I show it first for the imaging argument, and second for the resemblance one. Concerning imaging I come back to Lewis' proof of the fact that probabilities of Stalnaker conditionals are probabilities of their consequents under imaging on their antecedents. More precisely, I notice that the proof is essentially about probabilities of worlds and thus heavily relies on the connection between probabilities of worlds and probabilities of propositions. But what I have taken to be the propensity equivalents of world – that is physical systems – do not *have* probabilities, but rather *determine* probabilities. Moreover, although it could be imagined to attribute them probabilities (relative to a still more fundamental conditioning element), there is no reason why the relationship between these probabilities and probabilities of events should be analogous to the relationship between probabilities of worlds and probabilities of propositions in the subjectivist framework. As a consequence, Lewis' proof is not available but under a subjectivist conception of probability; the imaging argument falls. What is more is that the resemblance argument falls with it. Indeed, closer examination reveals that the resemblance stems from the confusion probabilities *in* worlds and probabilities *of* worlds. In the framework adopted by both Lewis and Stalnaker, worlds *have* probabilities but they do not *determine* probabilities. On the contrary, the propensity theory cannot be separated from the idea according to which physical systems

[52]The reference to time can be skipped since it does not vary with conditionalization under the proposed interpretation.

[53]Katsuno and Mendelzon (1992) p. 183.

determine probabilities.

B.6.4.2 Conditional probabilities?

As a consequence of what precedes, I cannot avoid to examine whether the interpretation I propose actually accounts for the properties of Bayesian conditionalization. Here, my first concern is about the evaluation of inverse conditional propensities. As far as Humphreys' example is concerned, I claim that the interpretation I propose leads to the same evaluation as the one advocated by McCurdy. More generally, the proposition seems to lead neither to one of the principles that Humphreys showed to be problematic, nor even to any principle to which could be opposed a new formal version of the paradox. I conclude that the proposition is not liable to Humphreys' paradox as such.

Second I envisage the question of the admissibility of the proposed interpretation. As in the absolute case, the question does not have a empirical question since there still does not exist a procedure enabling to measure propensities, that is to evaluate probabilities. Moreover, I claim that no argument in the type of Lewis' argument for absolute probabilities can be proposed. Indeed, such an argument would resort to twice conditional probabilities, which does not have any clear meaning. All this leaves only the possibility of assuming that the proposed interpretation accounts for the properties of Bayesian conditionalization. Exactly in the same way as I and most propensity proponents assume that the propensity theory is an interpretation of absolute probabilities, I assume that I have constructed an interpretation of Bayesian conditionalization.

This defence may seem weak. However, I suggest what the contribution of the section to the debate concerning the propensity interpretation of probabilities reside at least as much in the constructivist way the question is approached, as in the content of the proposition that is actually formulated. Moreover, the distinction between the approach and the proposition is taken into account in the coming discussion of the relationships between actual causation and probability given a propensity interpretation.

B.6.5 Actual causation and probabilities

B.6.5.1 Actual causation and conditional probabilities

In this paragraph I examine the consequences of the interpretation I propose as regards the relationship between causality and conditional probabilities. In other words, I come back to the very question that is tackled through

probabilistic theories of causality. The analyses in this paragraph depend on the content of the proposed interpretation.

First I wonder what the seminal idea on which probabilistic theories of causality are grounded becomes under the interpretation I propose for conditionalization. According to this idea, a cause may be characterized by its raising the probabilities of its effects. In the singular case, it is also fundamental to stipulate that both cause and effect occur. In other words, in the singular case, the seminal idea may be formulated as follows:

Proposition B.6 (Seminal idea in the singular case) *A causes B if and only if:*

1. *(a) A occurs;*
   (b) B occurs;
2. $Pr(B|A) > Pr(B|\overline{A})$.

In order to explicate this in terms of the proposed interpretation of conditionalization, one has to decide how Pr is defined. In other words, one has to decide by which fundamental conditional set of physical conditions it is determined. The question appears to be mainly that of the time at which this set is defined. My claim is that this time should be the one that is attached to event A, t_A. As a consequence, the proposed interpretation leads to the following reformulation of the seminal idea:

Proposition B.7 *A causes B if and only if:*

1. *(a) A occurs;*
   (b) B occurs;
2. *at t_A, the probability of B is greater in the system that is the most similar to the one under consideration among those that give A probability 1, than in the system that is the most similar to the one under consideration among those that give A probability 0.*

I explain why this proposition can be reformulated as follows:

Proposition B.8 (Seminal idea interpreted) *A causes B if and only if:*

1. *(a) A occurs;*
   (b) B occurs;
2. *at t_A, the probability of B is greater in the system under consideration, than it is in the system the most similar to it among those in which A does not occur.*

This proposition bears a strong resemblance to Lewis' analysis of what he calls "indeterministic causation" in Lewis (1986b). Indeed, Lewis' proposition is as follows:

Proposition B.9 (Lewis' analysis) *A causes B if and only if:*

1. *(a) A occurs;*
   (b) B occurs;
2. *the probability of B is greater in the actual world than it is in the world the most similar to the actual world amon those in which A does not occur,*

with the mentioned probability being explicitly considered a "single - case chance"[54]. The resemblance, still, calls for a more cautious comparison:
1. concerning the objects of the analyses, they are not exactly the same. First Lewis' analysis is an analysis of indeterministic causation whereas the seminal idea is meant to enable to characterize causation in general. Still it is clear from Lewis (1973) that what Lewis means by "indeterministic causation" is nothing but causation in general under the hypothesis that causes do not always necessitate their effects.[55] In other words, Lewis' indeterministic causation is causation (in general) under the hypothesis that motivates probabilistic theories of causality. The first difference between Lewis' analysis and the seminal idea under my interpretation vanishes. Second, under B.8, causality is clearly relative to a physical system, where Lewis' analysis deals with causality full stop. In a way that is analogous to the one that I took in subsection B.2.3, I will consider that causality (full stop) is causality relative to the actual world considered as a system. This allows me to compare the content of the analyses:
2. as far as the content of the analyses is concerned, I identify only one substantial difference: the seminal idea relies on the unqualified notion of probability-raising whereas, for Lewis, A causes B only if B would have been "*much less* probable"[56] if A had not occurred. As a consequence of this difference, Lewis' analysis is both more demanding and more vague than the seminal idea under the interpretation of conditionalization that I propose. In total, I claim that the proposed interpretation of conditionalization makes the seminal idea very similar to Lewis' analysis of indeterministic causation. This contributes to throw light on the question of the relationship between actual causation and conditional probabilities.

[54]Lewis (1986) p. 177.
[55]Lewis (1973) p. 559.
[56]Lewis (1986b) p. 177.

B.6.5.2 Actual causation and propensity theory

What I have exposed in the last paragraph makes sense only in relation to the content of my proposition for interpreting conditionalization. I now come to conclusions that do not rely on this content. These conclusions hinge on the following thesis: the discussion of Humphreys' paradox induces to distinguish between two concepts of causality.

To begin with, the relationship whose causal interpretation is central to the propensity theory is the relationship between sets of physical conditions and singular events. I have shown that Humphreys' paradox relies on the notion that the relationships between the conditioning event and the conditioned event of a conditional probability should be understood in a similar way. Correlatively, I have considered that Humphreys' paradox can be solved only if one considers that there is nothing causal in the propensity picture apart from the relationship between sets of physical conditions and singular events.

This does not entail that the propensity theory is incompatible with the usual idea according to which actual causation is a relation between singular events. More precisely, it appeared in last paragraph that causation between singular events may be *analyzed* in the context of the propensity theory. Contrary to causation between sets of physical conditions on the one hand and singular events on the other hand, causation between events is not primitive in the propensity framework. To finish with, I suggest that the distinction which appears is suitably analyzed in terms of Hall's distinction[57] between causation as production and causation as dependence.

[57] Hall (2001).

Conclusion

Conclusion

In this work, I have treated questions that are raised by probabilistic theories of causality. As explained in the introduction, these questions are different depending on whether one is interested in generic or in singular causation.

As far as generic causation is concerned, I have treated epistemological questions. The idea to approach these questions through Bayesian networks has found full justification in B.2. On the one hand, it appeared that the criterion for causality that is conveyed by causal Bayes nets come straight to the field of probabilistic theories of causality. On the other hand, it was shown that the differences between this criterion and our best probabilistic theories of generic causation accounts for the possibility of causal inference. Apart from this, the main results of the first part of the thesis are the following ones:

- in the context of causal inference, the hypotheses conveyed by causal Bayes nets are substantial, meaning that they cannot be made true through a trick;
- Bayes nets causal inference is *always* under suspicion of giving incorrect results. As a consequence, Bayes nets algorithms cannot be used as all there is to causal inference, but rather should be integrated to our hypothetico-deductive procedures. I have explored the ways along which this could be done;
- it is false that whether a system is deterministic or not is of no relevance concerning its satisfying the causal Markov condition. Positively, I have defined a very precise sense in which determinism is more favorable than indeterminism for the causal Markov condition.

As to singular causation, it appeared that it still calls for a probabilistic theory. More generally, the relationship between actual causation and probabilities are still to be clarified. My contribution of this clarification started from an analysis of the propensity theory of probability. It went on with a discussion of the problem the propensity theory has with conditional probabilities. Finally it consists in the following:

- a proposition for a propensity interpretation of conditionalization. Under this proposition, the idea that is seminal to probabilistic theories of causality is very similar to Lewis' analysis of indeterministic causation;

- the idea according to which the causal notion that is essential to the propensity theory is different from the one that probabilistic theories aim to characterize. Causation is production on the one hand and dependence on the other hand.

Bibliographie

Anscombe G. E. M. (1981) : Causality and determination. Reproduit in Sosa E. et Tooley M. (1993), 88-104.

Aristote (Métaphysique) : *Métaphysique*. Trad. Tricot, Vrin « Bibliothèque des textes philosophiques », 1991, Paris.

Aristote (Physique) : *Physique*. Trad. P. Pellegrin, GF-Flammarion, 2002, Paris.

Arntzenius F. (2005) : Reichenbach's common cause principle. In *The Stanford encyclopedia of philosophy* (édition hiver 2005), E. N. Zalta (éd.), URL = <http ://plato.stanford.edu/entries/physics-Rpcc/>.

Aronson J. (1971) : On the grammar of « cause ». *Synthese*, 22, 414–430.

P. Asquith and P. Kitcher (1985) (éds.) : *Proceedings of the biennial meeting of the Philosophy of Science Association, 1984*.

Bickel P.J., Hammel E.A. et O'Connell J.W. (1975) : Sex bias in graduate admissions : Data From Berkeley. *Science*, 187, 398-404.

Blalock H. (1991) : Are there really constructive alternatives to causal modeling ?. *Sociological methodology*, 21, 325–335.

Bouveresse R. (1981) : *Karl Popper ou le rationalisme critique*. Vrin « Histoire de la philosophie », 1998, Paris.

Buchanan B. et Shortliffe E. (1984) (éds.) : *Rule-based expert systems : The MYCIN experiments of the Stanford heuristic programming project*. Addison-Wesley « Series in artificial intelligence ».

Carroll J. W. (1991) : Property-level causation ?. *Philosophical studies*, 63, 245–270.

Cartwright N. (1979) : Causal laws and effective strategies. *Nous*, 13, 419–437.

Cartwright N. (1989) : *Nature's capacities and their measurement*. Oxford University Press.

Cartwright N. (1999) : Causal diversity and the causal Markov condition. *Synthese*, 121, 3–27.

Cartwright N. (2001) : What is wrong with Bayes nets ?. *The Monist*, 84, 242–264.

Cartwright N. (2002) : Against modularity, the causal Markov condition and any link between the two : Comments on Hausman and Woodward. *The British journal for the philosophy of science*, 53, 411–453.

Chickering D. (1996) : Learning Bayesian networks is NP-complete. In Lenz D. et Fisher H. (1996), 121–130.

Clogg C. et Haritou A. (1997) : The regression method of causal inference and a dilemma confronting this method. In McKim V. et Turner S. (1997), 83–112.

Collins J., Hall E. et Paul L. (2001) (éds.) : *Causation and counterfactuals*. MIT Press, Cambridge (Massassuchets).

Corfield D. et Williamson J. (2001) (éds.) : *Foundations of Bayesianism*. Kluwer « Applied logic series ».

Dancy J. et Sosa E. (1992) (éds.) : *A Companion to Epistemology*. Blackwell, « Blackwell companions to philosophy ».

Davis W. (1988) : Probabilistic theories of causation. In Fetzer (1988), 133–160.

Davis W. (1993) : Review of *Probabilistic causality* by Ellery Eells. *The philosophical review*, 102 : 3, 410–413.

de Finetti B. (1937) : La prévision, ses lois logiques, ses sources subjectives. *Annales de l'institut Henri Poincaré*, VII : 1, Gauthier-Villars, Paris.

Dowe P. (1992a) : Process causality and asymmetry. *Erkenntnis*, 37 : 2, 179–196.

Dowe P. (1992b) : Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of science*, 59, 195–216.

Drouet I. (2007) : Causal inference : How can Bayes nets contribute ? In Russo F. et Williamson J. (2007), 487–502.

Dupré J. (1992) : Review of *Probabilistic causality* by Ellery Eells. *Isis*, 83 :3, 528–529.

Durkheim E. (1895) : *Les règles de la méthode sociologique*. PUF « Quadrige », 2005, Paris.

Eagle A. (2004) : Twenty-one arguments against propensity analyses of probability. *Erkenntnis*, 60, 371–416.

Earman J. (1986) : *A primer on determinism*, chapitre I et II. Reidel, Dordrecht.

Eells E. et Sober E. (1983) : Probabilistic causality and the question of transitivity. *Philosophy of science*, 50 : 1, 35–57.

Eells E. (1991) : *Probabilistic causality*. Cambridge University Press « Cambridge studies in probability, induction and decision theory ».

Esposito C. et Porro P. (2002) (éds.) : *Quaestio-Annuario di storia della metafisica* vol. 2. Brepols, Turnhout.

Fair D. (1979) : Causation and the flow of energy. *Erkenntnis*, 14, 219–250.

Felouzis G. (2003) : La ségrégation ethnique au collège et ses conséquences. *Revue française de sociologie*, 44 : 3, 413–447.

Fetzer J. (1970) : Dispositional probabilities. *PSA : Proceedings of the biennial meeting for the Philosophy of Science Association*, 473–482.

Fetzer J. (1981) : *Scientific knowledge : Causation, explanation, and corroboration*. Dordrecht « Boston studies in the philosophy of science ».

Fetzer J. (1982) : Probabilistic explanations. *PSA : Proceedings of the biennial meeting for the philosophy of science association*, vol.2, 194–207.

Fetzer J. (1988) (éd.) : *Probability and causality. Essays in honor of Wesley C. Salmon*. Reidel, Dordrecht.

Freedman D. (1987) : As others see us : A case study in path analysis. *Journal of educational statistics*, 12, 101–128.

Freedman D. (1991) : Statistical models and leather shoe. *Sociological methodology*, 21, 291–313.

Freedman D. et Humphreys P. (1999) : Are there Algorithms that discover causal Structure ?. *Synthese*, 121, 29–54.

Galavotti M.C., Suppes P. et Costantini D. (2001) : *Stochastic causality*. CSLI publications, Stanford.

Gärdenfors P. (1992) (éd.) : *Belief revision*. Cambridge University Press.

Gillies D. (1972) : Operationalism. *Synthese*, 25, 1–24.

Gillies D. (2000a) : *Philosophical theories of probability*. Routledge, Londres et New-York.

Gillies D. (2000b) : Varieties of propensity. *The British journal for the philosophy of science*, 51, 807–835.

Gillies D. (2002) : Causality, propensity, ans Bayesian networks. *Synthese*, 132, 63–88.

Good I. J. (1961a) : A Causal Calculus (I). *The British journal for the philosophy of science*, 11 : 44, 305–318.

Good I. J. (1961b) : A Causal Calculus (II). *The British journal for the philosophy of science*, 12 : 45, 43–51.

Glymour C., Scheines R. et Spirtes P. (1988) : Exploring causal structure with the tetrad program. *Sociological methodology*, 18, 411–448.

Hájek A. (2007) : Interpretations of probability. In *The Stanford encyclopedia of philosophy* (édition automne 2007), E. N. Zalta (éd.), URL = <http ://plato.stanford.edu/archives/fall2007/entries/probability-interpret/>.

Hall N. (2001) : Two concepts of causation. In Collins, Hall et Paul (2001), 225–274.

Halpern J. et Pearl J. (2005a) : Causes and explanations. A structural-model approach. Part I : Causes. *The British journal for the philosophy of science*, 56 : 4, 843–887.

Halpern J. et Pearl J. (2005a) : Causes and explanations. A structural-model approach. Part I : Explanations. *The British journal for the philosophy of science*, 56 : 4, 889–911.

Harman G. (1965) : The inference to the best explanation. *Philosophical review*, 74, 88–95.

Harman G. (1992) : Induction : Enumerative and Hypothetical. In J. Dancy and E. Sosa (1992), 200–206.

Hausman D. (1998) : *Causal asymmetries*. Cambridge University Press « Cambridge studies in probability, induction, and decision theory ».

Hausman D. and Woodward J. (1999) : Independence, invariance and the causal Markov condition. *The British journal for the philosophy of science*, 50, 521–583.

Hausman D. and Woodward J. (2004) : Modularity and the causal Markov condition : a restatement. *The British journal for the philosophy of science*, 55, 147–161.

Hesslow G. (1976) : Discussion : Two notes on the probabilistic approach to causality. *Philosophy of science*, 43, 290–292.

Hitchcock C. R. (1995) : The mishap at Reichenbach fall : Singular vs. general causation. *Philosophical studies*, 78 : 3, 257-291.

Hitchcock C. R. (2002) : Probabilistic causation. In *The Stanford encyclopedia of philosophy* (édition automne 2002), E. N. Zalta (éd.), URL = <http ://plato.stanford.edu/archives/fall2002/entries/causation-probabilistic/>.

Hume D. (1739) : *Traité de la nature humaine. Livre I : L'entendement*. Trad. P. Baranger et P. Saltel, GF-Flammarion, 1995, Paris.

Hume D. (1748) : *Enquête sur l'entendement humain*. Trad. A. Leroy, GF-Flammarion, 1983, Paris.

Humphreys P. (1985) : Why propensities cannot be probabilities. *The philosophical review*, 94 : 4, 557–570.

Humphreys P. (1986) : Causation in the social sciences : An overview. *Synthese*, 68 :1, 1–12.

Humphreys P. (1989) : *The chances of explanation : Causal explanation in the social, medical, and physical sciences*. Princeton University Press.

Humphreys P. (1994) (éd.) : *Probability and probabilistic causality, Patrick Suppes : Scientific philosopher* vol.1. Kluwer, Dordrecht.

Humphreys P. et Freedman D. (1996) : The grand leap. Review of Peter Spirtes, Clark Glymour, and Richard Scheines [1993] : *Causation, prediction, and search. The British journal for the philosophy of science*, 47, 113–123.

Humphreys P. (2004) : Some considerations on conditional chances. *The British journal for the philosophy of science*, 55, 667–680.

Jeffrey R. (1980) (éd.) : *Studies in inductive logic and probability* vol.II. University of California press.

Katsuno A. et Mendelzon A. (1992) : On the difference between updatig a knowledge base and revising it. In Gärdenfors (1992), 183–203.

Kenny D. (1979) : *Correlation and causality*. Wiley, New-York. Version révisée téléchargeable en ligne, URL = <http ://davidakenny.net/doc/cc_v1.pdf>.

Kistler M. (1999) : *Causalité et lois de la nature*. Vrin « Mathesis », Paris.

Kistler M. (2002) : Causation in contemporary analytical philosophy. In Esposito C. et Porro P. (2002), 635–668.

Kline R. (1998) : *Principles and practive of structural equation modeling*. Deuxième édition : Guilford Press, New-York, 2005.

Korb K. et Wallace C. (1997) : In search of the philosopher's stone : Remarks on Humphreys and Freedman's critique of causal discorvery. *The British journal for the philosophy of science*, 48, 543–553.

Körner S. et Pryce M. (1957) (éds.) : *Observation and interpretation. Proceedings of the ninth symposium of the Colston research society*. Colston Papers, University of Bristol.

Kwoh C. et Gillies D. (1996) : Using hidden nodes in Bayesian networks. *Artificial intelligence*, 88, 1–38.

Kyburg H. et Smokler H. (1980) (éds.) : *Studies in subjective probability* vol.II. Deuxième édition : Robert E. Krieger Publishing Company, Huntington (N.-Y.).

Kyburg H. (2002) : Don't take unnecessary chances !. *Synthese*, 132, 9–26.

Lauritzen S. et Spiegelhalter D. (1988) : Local computations with probabilities in graphical structures and their applications to expert systems (with discussion). *Journal of the royal statistical society series B*, 50 : 2, 157–224.

Lenz D. et Fisher H. (1996) (éds.) : *Learning from data*. Springer-Verlag « Lecture notes in statistic ».

Lewis D. (1973) : Causation. *The journal of philosophy*, 70, 556–567.

Lewis D. (1976) : Probabilities of conditionals and conditional probabilities. *The philosophical review*, 85 : 3, 297–315.

Lewis D. (1980) : A subjectivist's guide to objective chance. In Jeffrey R. (1980), 263–293.

Lewis D. (1986a) : *Philosophical papers*, volume II. Oxford university press.

Lewis D. (1986b) : Postscripts to « Causation ». In Lewis (1986a), 172–213.

Mackie J. (1974) : *The cement of the universe. A study of causation*. Oxford university press, « Clarendon library of logic and philosophy », 1980.

McCurdy C. : Humphreys' paradox and the interpretation of inverse conditional propensities. *Synthese*, 108, 105–120.

McKim V. et Turner S. (1997) (éds.) : *Causality in crisis ? Statistical methods and the search for causal knowledge in the social sciences*. University of Notre-Dame Press.

Mellor D. (1988) : On raising the chances of effects. In Fetzer (1988), 229-239.

Mill J. S. (1853) : *A system of logic, ratiocinative and inductive : being a connected view of the principles of evidence and the methods of scientific investigation*. Huitième édition : Harper and Brothers, 1874, New-York.

Miller D. (1994) : *Critical rationalism. A restatement and defence*. Open Court Publishing Company, Chicago et La Salle.

Miller D. (2002) : Propensities may satisfy Bayes' theorem. *Proceedings of the British academy*, 113, 111–116.

Milne P. (1986) : Can there be a realist single-case interpretation of probability ?. *Erkenntnis*, 25 :2, 129–132.

Neapolitan R. (1990) : *Probabilistic reasoning in expert systems : theory and algorithms*. Wiley, New-York.

Pearl J. (1988) : *Probabilistic reasoning in intelligent systems*. Morgan Kaufman, San Mateo (Californie).

Pearl J. (2000) : *Causality. Models, reasoning and inference*. Cambridge University Press.

Popper K. (1934) : *La logique de la découverte scientifique*. Trad. A. Thyssen – Rutten et P. Devaux, Payot « Bibliothèque scientifique », 1990, Paris.

Popper K. (1957) : The propensity interpretation of the calculus of probability. In Körner S. et Pryce M. (1957), 65–70 et 88–89.

Popper K. (1959) : The propensity interpretation of probability. *The British journal for the philosophy of science*, 10, 25–42.

Popper K. (1982a) : *La théorie quantique et le schisme en physique*. Trad. Dissaké, Hermann, 1996, Paris.

Popper K. (1982b) : *L'univers irrésolu*. Trad. Bouveresse, Hermann, 1984, Paris.

Popper K. (1983) : *Le réalisme et la science*. Trad. Boyer et Andler, Hermann, 1990, Paris.

Popper K. (1990) : *Un univers de propensions*. Trad. Boyer, L'Éclat « Tiré à part », 1992, Paris.

Quine W. (1960) : *Word and object*. MIT Press, Cambridge (Massassuchets).

Ramsey F. (1926) : Truth and probability. In Kyburg H. et Smokler H. (1980), 23–52.

Reichenbach H. (1956) : *The direction of time*. University of California Press, Berkeley et Los Angeles.

Rescher N. (1968) (éd.) : *Studies in logical theory*. Blackwell « American philosophical quaterly monograph series », 2.

Rescher N. (1969) (éd.) : *Essays in honour of Carl G. Hempel*. Reidel, Dordrecht.

Rosen D. (1978) : In defence of a probabilistic theory of causality. *Philosophy of Science*, 45, 604–613.

Russo F. (2005) : *Measuring variations. An epistemological account of causality and causal modelling*. Thèse de doctorat, université catholique de Louvain, institut supérieur de philosophie.

Russo F. et Williamson J. (2007) : *Causality and probability in the sciences*. College publications « Text in philosophy series ».

Salmon W. (1966) : *The foundations of scientific inference*. University of Pittsburgh Press.

Salmon W. (1979) : Propensities : A discussion review. *Erkenntnis*, 14, 183–216.

Salmon W. (1980) : Probabilistic causality. Repris dans Sosa E. et Tooley M. (1993), 137–153.

Salmon W. (1984) : *Scientific explanation and the causal structure of the world*. Princeton University Press.

Shimony A. (1988) : An adamite derivation of the calculus of probability. In Fetzer J. (1988), 151–161.

Shortliffe E. et Buchanan B. (1984) : A model of inexact reasoning in medicine. In Buchanan B. et Shortliffe E. (1984), 233–262.

Skyrms B. (1980) : *Causal necessity*. Yale University Press, New-Haven et Londres.

Skyrms B. (2004) : *The stag hunt and the evolution of social structure*. Cambridge University Press.

Sober E. (1985) : Two concepts of cause. In P. Asquith and P. Kitcher (1985), vol. 2, 405-424.

Sober E. (1986) : Causal factors, causal inference, causal explanation. *Aristotelian society supplementary volume*, 40, 97–113.

Sober E. (1988) : The principle of the common cause. In Fetzer (1988), 211–228.

Sosa E. et Tooley M. (1993) (éds.) : *Causation*. Oxford University Press.

Spirtes P., Glymour C., Scheines R. (1990) : Simulation studies of the reliability of computer aided specification using TETRAD II, EQS, and LISREL programs. *Sociological methods and research*, 19, 3–66.

Spirtes P., Glymour C., Scheines R. (1991) : An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, 9, 62–72.

Spirtes P., Glymour C., Scheines R. (1993) : *Causation, prediction and search*. Deuxième édition : Springer-Verlag, MIT Press, 2001.

Spirtes P., Glymour C., Scheines R. (1997) : Reply to Humphreys and Freedman's review of *Causation, prediction, and search. The British journal for the philosophy of science*, 48 : 4, 555–568.

Spohn W. (1994) : On the properties of conditional independence. In Humphreys P. (1994), 173-194.

Spohn W. (2001) : Bayesian nets are all there is to causal dependence. In Galavotti M.C., Suppes P. et Costantini D. (2001), 157–172.

Stalnaker R. (1968) : A theory of conditionals. In Rescher N. (1968), 98–112.

Steel D. (2005) : Indeterminism and the causal Markov condition. *The British journal for the philosophy of science*, 56 : 1, 3–26.

Suppes P. (1970) : *A probabilistic theory of causality*. North Holland Publishing Company, Amsterdam.

Teller P. (1973) : Conditionalization and observation. *Synthese*, 26 : 2, 218–258.

Vickers J. (2006) : The problem of induction. In *The Stanford encyclopedia of philosophy* (édition hiver 2006), E. N. Zalta (éd.), URL = <http ://plato.stanford.edu/archives/win2006/entries/induction-problem/>.

Walliser B. et Zwirn D. (2002) : Can Bayes' rule be justified by cognitive rationality principles ?. *Theory and decision*, 53, 95–135.

Williamson J. (2001a) : Foundations for Bayesian networks. In Corfield D. and Williamson J. (2001).

Williamson J. (2002) : Learning causal relationships. Technical report 02/02, London School of Economics, Center for the Philosophy of Natural and Social Sciences.

Williamson J. (2005) : *Bayesian nets and causality*. Oxford University Press, New-York.

Woodward J. (1994) : Paul Humphreys [1989] *The chances of explanation* (Book review). *The British journal for the philosophy of science*, 45 : 1, 353–374.

Wright S. (1921) : Correlation and causation. *Journal of agricultural research*, 20, 557–585.

Wright S. (1934) : The method of path coefficients. *Annals of mathematical statistics*, 5 :3, 161–215.

Yule G. (1903) : Notes on the theory of association of attributes in statistics. *Biometrika*, 2, 121–134.

Index des noms

Index des notions

Causalité et probabilités : réseaux bayésiens, propensionnisme

Résumé : Les théories probabilistes de la causalité apparaissent dans les années 1960 corrélativement de la critique de l'idée selon laquelle la causalité serait une relation de nécessitation. Le présent travail traite de questions soulevées par l'état actuel du développement de ces théories. En ce qui concerne la causalité générique, on peut considérer que l'analyse conceptuelle de ses rapports avec les probabilités est achevée. Les questions qui se posent aujourd'hui sont donc épistémologiques. Plus exactement, les questions traitées dans ce travail portent sur l'inférence aux causes génériques en tant qu'elle est fondée sur les réseaux bayésiens causaux. De façon sensiblement différente, la question du rapport entre la causalité singulière et les probabilités n'est pas complètement réglée du point de vue conceptuel. Nous abordons cette question à partir d'une analyse de la relation entre la causalité et la théorie propensionniste des probabilités.

Causality and probability: Bayesian networks, propensities

Summary: Probabilistic theories of causality emerged in the 1960s together with the criticism of the idea that causation is a relation of necessitation. The present work treats questions that are raised by the current state of development of the field of probabilistic theories of causality. Concerning generic causation, the relationship between causality and probability can be considered conceptually analyzed and open questions are epistemological. More exactly, I am interested in inference to generic causes as it is grounded on causal Bayesian networks. By contrast, the conceptual question of the relationship between probabilities and singular causation is not completely answered. This question is broached through an analysis of the relationship between causality and the propensity theory of probability.

Discipline : Philosophie

Mots-clés : Philosophie des sciences ; Philosophie des probabilités ; Épistémologie ; Causalité ; Théories probabilistes de la causalité ; Réseaux bayésiens ; Propensionnisme ; Philosophie des sciences sociales.

Équipe d'accueil : Institut d'Histoire et de Philosophie des Sciences
et des Techniques (UMR 8590)
13, rue du Four
75006, Paris.